

# nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE



## **BODY ELECTRIC**

Molecular basis of sensory  
discrimination in sharks  
and skates **PAGE 122**

### **EQUALITY**

#### **DIVERSITY IN THE LAB**

Why inclusive teams  
excel at science

**PAGES 19 & 149**

### **GENETICS**

#### **BLOOD PRODUCTS**

A genomic atlas of the  
proteins in human plasma

**PAGE 73**

### **PLANETARY SCIENCE**

#### **LIGHTNING ON JUPITER**

Broadband emissions reveal  
nature of Jovian discharges

**PAGE 87**

**NATURE.COM/NATURE**

7 June 2018

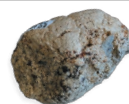
Vol. 558, No. 7708

# THIS WEEK

## EDITORIALS

**AGRICULTURE** Europe's advisers offer sensible measures on crop protection **p.6**

**WORLD VIEW** Chew on better ways to measure food production **p.7**



**FOSSILS** Poo shows ancient dogs had bone-crunching diets **p.9**

## Science benefits from diversity

*Improving the participation of under-represented groups is not just fairer — it could produce better research.*

**L**ab groups, departments, universities and national funders should encourage participation in science from as many sectors of the population as possible. It's the right thing to do — both morally and to help build a sustainable future for research that truly represents society.

A more representative workforce is more likely to pursue questions and problems that go beyond the narrow slice of humanity that much of science (biomedical science in particular) is currently set up to serve. Widening the focus is essential if publicly funded research is to protect and preserve its mandate to work to improve society. For example, a high proportion of the research that comes out of the Western world uses tissue and blood from white individuals to screen drugs and therapies for a diverse population. Yet it is well known that people from different ethnic groups can have different susceptibility to some diseases.

Many people are working to improve diversity in science and the scientific workforce. Some have been trying hard for decades, but not all are succeeding. This week, *Nature* highlights examples of success from across the world. They are inspiring, and show what can — and must — be done.

To boost recruitment and participation in science among some under-represented groups is difficult. Statistics from the US National Science Foundation show that the representation of minority ethnic groups in the sciences would need to more than double to match the groups' overall share of the US population.

As we highlight in a Careers piece this week (page 149), there are steps that groups, departments and institutions can take to try to draw from a broader pool of talent. Some of these demand effort to reach out to under-represented communities, to encourage teenagers who might otherwise not consider science as an option. Even the wording of job advertisements can put people off — candidates from some backgrounds might be less likely to consider themselves 'outstanding' or 'excellent', and so might not even apply. Yet diversity efforts should not stop when people are through the door. To retain is as important as to recruit — mentoring and support is essential for all young scientists, and especially so for those who have been marginalized by academic culture.

Projects to boost participation are often the passion and work of a few dedicated individuals. More institutions and funders should seek, highlight and support both the actions and the individuals.

There are moral and ethical reasons for institutions to act. And there are other potential benefits, too. Firms are recognizing that diversity — and associated attitudes and behaviours — is a business issue. A report from consultancy firm McKinsey earlier this year was just the latest to set out the healthy relationship between a company's approach to inclusion and diversity and its bottom line. The report, *Delivering through Diversity*, reaffirms the positive link between a firm's financial performance and its diversity — which it defines in terms of the proportion of women and the ethnic and cultural composition of the leadership of large companies.

Could something similar be true in science? As we discuss in a News Feature this week (page 19), some studies suggest that a team with a good mix of perspectives is associated with increased productivity.

**"The lack of diversity in science is everyone's problem."**

Concerted action to effect change on recruitment and retention can and does make a difference (see T. Hodapp and E. Brown *Nature* 557, 629–632; 2018). More effort across the board is overdue. The lack of diversity in science is everyone's problem.

Everyone has a responsibility to look around them, to see the problem for what it is, and to act — not just to assume it is someone else's job to fix it. ■

## Targeting cancer

*Cancer treatments tailored to individual tumours must not be oversold.*

**C**ancer specialist Leonard Saltz received a letter earlier this year from someone who had watched a television programme about the promise of 'precision oncology'. A patient had taken a few pills and seen his tumour disappear, the letter said. Could the same be done for his sick father?

Saltz, who works at Memorial Sloan Kettering Cancer Center in New York City, was distressed. "That's what people think precision oncology is," he says. "And, gosh, I wish that were so."

It's not unusual for the promise and perception of new cancer treatments to run ahead of the reality. And it's true that precision oncology is promising. The practice — which relies on finding weak spots in a particular tumour's genetic make-up that can be targeted by drugs — is growing, and new results feature strongly this week at the annual meeting of the American Society of Clinical Oncology in Chicago, Illinois — cancer medicine's biggest annual meeting. But talk of potential benefits must be tempered by clinical reality.

Over the past decade, advances in genomic sequencing and analysis have yielded a steady stream of information about the genetic mutations that can drive cancer. The studies have revealed that even cancers of the same type, such as breast tumours, can be very different genetically. From that has grown the hope that drugs can be tailored to a tumour's genetic anomalies, resulting in a treatment with, ideally, fewer side effects and greater efficacy than conventional therapies. A handful of such drugs are already on the market. One, Herceptin (trastuzumab), has already increased survival rates for women with particular types of breast cancer.

This model of precision oncology is now at a turning point, as some

of the long-anticipated changes to cancer care work their way from bench to bedside — ones that would allow precision oncology to be scaled up. In the past year, the US Food and Drug Administration has issued its first approval of a genetic test that can detect mutations in hundreds of cancer-associated genes. Also a first, the agency approved a drug for the treatment of any solid tumour bearing a particular genetic signature, regardless of what tissue the tumour originated in.

Health services around the world are talking up the role of DNA and genomics in a new era of personalized medicine. But the utility of increasingly expensive cancer tests and medications that will help only a minority of patients is also being fiercely debated. Some 30 or so cancer drugs have so far been linked to a specific genetic signature. Many people have benefited, but some will relapse later as their tumours become resistant to the therapy.

Against this backdrop, clinicians are left facing ill people and trying to work out what to do. Whose tumours should be sequenced, and when? How often should one patient's tumour be sequenced? What kind of sequencing should be done — 50 genes, 400 genes, a full genome? How should physicians interpret genetic variants and conflicting data?

And over it all hangs the painful question that health-care systems everywhere must grapple with: at what point does the potential for benefit outweigh the cost of sequencing and the treatment that follows?

Researchers can help to pave precision oncology's path to the clinic. More research on cancer genetics might reveal roles for

as-yet-unexplained genetic variants. Such studies would also help researchers to unpick the effects of combinations of genetic variants, a consideration that is likely to become more important as clinicians sequence larger sets of a tumour's genes. Also useful is the growing emphasis in cancer research on testing targeted therapies in combination with one another, and together with drugs that provoke immune responses to cancer. From a clinical perspective, better and more-thorough screening should identify the people most likely to benefit.

**“Clinicians are left facing ill patients and trying to work out what to do.”**

Precision oncology increases the range of treatment options — but so far for only a relatively small number of people. Yet clinicians say that media reports of miracle cures have painted a much rosier picture, fuelled by anecdotes about exceptional responders who experience dramatic, but highly unusual, responses to treatment. In the United States, the problem is compounded by advertisements — from pharmaceutical companies and treatment centres — aimed directly at people with cancer. Enthusiasm for the possibilities of precision oncology has led too many involved to present the option with too much optimism. By its very nature, each precision cancer drug is destined to help only a fraction of people. Everyone with cancer wants, understandably, to be in that fraction. Hope is important. But all parties need to be sensitive to how the promise of precision medicine is communicated to patients — and to their physicians. ■

## Food chain

*European advisers set out a path to a sustainable future for food production.*

When Europe scrapped its chief scientific adviser role and instead installed a committee of experts in 2016, there were questions about how well the system would function. Very well indeed, is the answer — at least if a report released by the expert group on 4 June is anything to go by.

Ostensibly, the opinion document from the European Union's Group of Chief Scientific Advisors discusses how the EU authorizes plant protection products (chiefly insecticides and herbicides). But it goes further, offering sound advice on how to reform aspects of the EU's infamous bureaucracy and convoluted decision-making mechanisms for agriculture. And written between the lines is a clear and simple message, which Europe needs to take on board sooner rather than later: that the region's approach to food production is fragmented and hopelessly unsuited to future needs.

The report is the latest in a series of papers by the group, all “from a scientific point of view”. It will feed into specific discussions about, for example, how the commission can better integrate the functions of its agriculture, food, environment and research directorates. That is important if Europe is to set out a coherent plan for a sustainable future. At present, it is too easy for policy-making on a continent-wide level to be paralysed, as seen with research into and applications for genetically modified (GM) organisms. And, as shown by a controversy late last year over the approval of the herbicide glyphosate by the EU, there is insufficient public trust in the process.

The committee was tasked by the European Commission (EC) to work out whether the current system for approval of these products could be more effective, efficient and transparent. The report makes some sensible suggestions for improving transparency, some of which can and should be implemented quickly in the existing approval process. It recommends a new public IT platform to store the relevant data, case studies and information on cultural and historical

differences in agricultural practice that need to be built into models that assess risk. It calls for more systematic updates to the assessment of active substances when new data become available. It supports more monitoring and analysis of how pesticides and herbicides accumulate in the environment and in wildlife. And it suggests that mandatory pre-registration of the lab studies that companies will rely on to show their chemical is safe (including the lab location, the types of test planned and what will be learnt from them) would help to address concerns about the independence and objectivity of industry-sponsored studies.

More fundamentally, the report suggests some structural and systemic changes to the approval process. These range from clarifying levels of acceptable risk (current regulations invoke the precautionary principle to demand no harm to health or the environment, which is unachievable in practice), to recognizing that taking no action (for example, not applying a pesticide) also carries risks. Furthermore, the report recommends bringing the risk-assessment process within the control of the EC (it is currently outsourced to member states).

These types of change are more difficult to implement — not least because, at present, nations have control over the process (and, in the GM case, a de facto veto). National politicians will not surrender that control lightly, particularly in countries such as Germany, where anti-GM feeling has huge influence.

The particular wisdom of the latest report is in its recognition that, for such political changes to become possible, the focus of the public debate must shift from single issues in agriculture to the bigger question for society — how do we want to create sustainable agriculture in Europe and ensure quality food production, and how much are we prepared to pay for it? Pesticides and herbicides have a part to play, but so do complex and sometimes conflicting issues that have a relationship to agriculture: fertilizers, food chains and environmental protection in general. Tighter controls of pesticides, for example, will affect these other aspects and have costs and benefits to society. Such a discussion will go beyond a strictly scientific point of view, and must account for values and human judgements.

A good start would be for the commission to arrange a high-profile workshop for all relevant parties — including the public, non-governmental organizations, scientists and companies — to kick-start the process. Good advice alone is not enough. ■



## Smarter metrics will help fix our food system

Think less about bigger crop yields, and more about better lives, says Pavan Sukhdev, as more-comprehensive evaluation techniques are unveiled.

Today's food systems are broken. Our diets are the leading cause of disease. Some 800 million people worldwide still suffer from hunger, while more than 2 billion are overweight or obese. As much as 57% of global greenhouse-gas emissions come from food-related activities, which include everything from clearing land for agriculture, to growing, gathering, processing and packaging, to transporting farm goods and disposing of waste.

I never fail to be astonished at the inadequacy of the metrics we use to evaluate these systems. The most common yardstick is 'productivity per hectare'. This measure of the yield or value of a particular crop relative to the area of the land on which it was grown is too narrow. We need alternatives that account for the interacting complex of agricultural lands, pastures, inland fisheries, natural ecosystems, labour, infrastructure, technology, policies, markets and traditions that are involved in growing, processing, distributing and consuming food.

We've seen benefits from broader metrics elsewhere. Health experts know to look beyond calorie counts to understand nutrition. Policy-makers are less willing to accept gross domestic product as a proxy for national well-being and are turning to expanded measures of progress. And some private-sector leaders are looking beyond financial profit and loss, and assessing the impacts of their business on natural, human and social capital.

At last, after 4 years of work involving more than 150 people, including myself, there is a framework and methodologies for more-comprehensive food metrics. The effort has culminated in a report released this week by the United Nations Environment Programme called 'The Economics of Ecosystems and Biodiversity for Agriculture and Food' (TEEBAgriFood). It demonstrates how to capture the complex reality of food systems through a wide-angle lens. If this work helps to divert even a fraction of brain power and political will from maximizing yields to maximizing broader benefits, it will make for healthier people, communities and ecosystems.

TEEBAgriFood sets out an evaluation approach that accounts for the impacts of the food system on livelihoods, equity, food security, health, greenhouse-gas emissions, water quality and biodiversity. This approach can reveal effects that are invisible using assessments that consider only the production and marketing segments of food-value chains. The insights gained can support better decision-making for policymakers, farmers, agribusinesses and civil society.

For instance, one study based in New Zealand (H. S. Sandhu *et al. Ecol. Econ.* 64, 835–848; 2008) used a broader framework to compare conventional and organic agriculture, and found that important, non-marketed, ecosystem services have much higher value in the organic sector. Researchers considered the benefits provided by 15 conventional and 14 organic fields used for crops such as carrots,

peas and wheat. These benefits included two 'provisioning' ecosystem services (food and raw materials) and nine 'regulating and supporting' services, such as pollination, biological pest control and nutrient cycling. Organic farming practices such as composting and maintaining vegetation cover lead to higher biomass and diversity, below and above ground. Conventional agriculture suppresses these and diminishes soil health, farm biodiversity, water quality and air quality. The study found that the total economic value of ecosystem services from organic fields ranged from US\$1,610 to US\$19,420 per hectare per year; that from conventional fields ranged from \$1,270 to \$14,570 per hectare per year.

This analysis only partially employed the TEEBAgriFood framework because it covered only production. To investigate other trade-offs and impacts, researchers should also compare food affordability and the impacts of nutrition, human health and social equity between the two agricultural systems.

A second example concerns pesticide policies. In the late 1980s, Thailand began encouraging the use of pesticides to increase agricultural yields. In 2010, productivity gains started to fall and policymakers became increasingly aware of pesticides' harmful effects on the environment and health. Researchers examined the effects of increasing taxes to make pesticides more expensive, and of encouraging farmers to adopt non-chemical forms of pest management (S. Praneetvatakul *et al. Environ. Sci. Policy* 27, 103–113; 2013). They considered the costs of enforcing food-safety standards. They also examined the risks of exposure to chemical agents. These risks were higher for farm workers than for consumers, so the researchers argued for an increased environmental tax. This, combined with support to encourage a switch to new farming practices, would deliver the greatest benefits most effectively, the researchers argued. Standard productivity measures could not have helped to assess such nuanced effects.

We need many more studies to show how considering broad impacts leads to conclusions that differ from those based simply on market prices of output. Several pilots are planned or under way, and I encourage more researchers to test the evaluation tool in studies of farming, food products and policy scenarios, as well as in dietary comparisons. If we can keep the pressure of evidence strong for just five years, I expect to start to see large changes in how agricultural, health and environmental ministries across the world set policies, incentives, subsidies and taxes.

Only if we diagnose our food system honestly, can we heal it. ■

Pavan Sukhdev is founder and chief executive of GIST Advisory, a sustainability consultancy based in Mumbai, India.  
e-mail: pavan@gistadvisory.com

ONLY IF WE  
DIAGNOSE  
OUR FOOD SYSTEM  
HONESTLY,  
CAN WE  
HEAL IT.

# SEVEN DAYS

The news in brief

## SPACE

### Closer look at Ceres

On 31 May, NASA's Dawn spacecraft began moving into its closest orbit yet around the dwarf planet Ceres. Dawn fired its ion thrusters to nudge itself into a trajectory that will take it as close as 35 kilometres above the planet's surface. The spacecraft will target features such as the Occator Crater — in which nestle bright patches of salt deposits — to study the geology. Scientists also plan to use the new orbit to collect images of Ceres and to investigate its composition. The fly-by will take Dawn ten times closer to the dwarf planet, which is also the Solar System's largest asteroid, than it has ever been.

## RESEARCH

### Saltwater rice

Chinese scientists have started large field trials of rice designed to grow in salty environments. If the experiments are successful, the hybrid rice could boost the country's crop yield. On 28 May, a team led by rice breeder Yuan Longping from the China National Hybrid Rice Research and Development Center in Changsha started planting

#### MENTORING AWARDS

This year's *Nature* Awards for Mentoring in Science will recognize achievements in the south of the United States, as defined by the US Census Bureau (see [go.nature.com/2kopl0s](http://go.nature.com/2kopl0s)). For a description of the awards, see [go.nature.com/2jfrjbx](http://go.nature.com/2jfrjbx). For details of how to nominate people for this year's competition, see [go.nature.com/2hkl4go](http://go.nature.com/2hkl4go). The deadline for nominations is Monday 30 July.



JOHAN ORDONEZ/AFP/GETTY

## Guatemala volcano wreaks havoc

The Fuego volcano in Guatemala erupted on 3 June, sending a superheated avalanche of rock and gas downhill that has killed at least 69 people. It was the deadliest eruption in Guatemala in more than a century. Volcanic

ash rose some 6 kilometres into the air, and fell on Guatemala City about 40 kilometres away. Rain falling on the ash created mudflows that destroyed at least one bridge. Fuego is one of Latin America's most active volcanoes.

176 varieties of hybrid rice at sites across China. The varieties have been bred to grow in tidal flats and other salt-rich environments. The team hopes to find a strain that can be planted on the roughly 7 million hectares of land that is too salty for current strains of rice. But some scientists are sceptical about whether the 'saltwater' rice will be able to grow in these conditions.

### Cancer-drug boom

More than 1,100 cancer drugs and vaccines are in clinical trials or up for evaluation by the US Food and Drug Administration, according to a report by the Pharmaceutical Research and Manufacturers of America (PhRMA). This is up from roughly 830 such therapies in 2015. The report, released by the lobbying organization

on 30 May, highlights the intense activity in the sector (see [go.nature.com/2j2gzg9](http://go.nature.com/2j2gzg9)). US pharmaceutical companies are testing hundreds of drugs and vaccines aimed at leukaemia, lymphoma and lung cancers alone, and more than 200 others target breast cancer, brain tumours and skin cancers.

## EVENTS

### Hurricane toll

Nearly 5,000 people in Puerto Rico may have died as a result of Hurricane Maria, which hit the island late in 2017 — more than 70 times the official government estimate. In a study published on 29 May, researchers surveyed a random selection of 3,300 homes, and found that 38 people had died between 20 September — the date that Hurricane Maria

reached Puerto Rico — and 31 December (N. Kishore *et al.* *N. Engl. J. Med.* <http://doi.org/cqgr>; 2018). When they extrapolated to account for Puerto Rico's population, they concluded that the national death rate was probably 62% higher than for the same period in 2016. The researchers say this figure is likely to be an underestimate, although it dwarfs the Puerto Rican government's official record of 64 deaths related to Maria. The researchers believe that many of the deaths were a result of disrupted medical services.

## POLICY

### Student boost

The European Commission has proposed opening up Erasmus+, the European Union's student-exchange programme, to all countries.

KYODO NEWS VIA GETTY

Erasmus+ enables university students, including PhD researchers, to study abroad. The plan would allow the United Kingdom to participate in the programme after Brexit, as well as researchers worldwide to gain experience at a university in Europe. The commission's proposal, published on 30 May, also recommends doubling the Erasmus+ budget to €30 billion (US\$35.2 billion) for the programme's next instalment, which will run from 2021 to 2027. The boost would fund the participation of around 12 million people, up from 4 million in the current round. The plan must now be approved by the European Parliament and the EU Council of Ministers.

## ENVIRONMENT

## Whale hunt

Japan's latest annual Antarctic whale hunt — which the country says is for scientific purposes — killed 333 minke whales between 8 December 2017 and 28 February 2018. The International Whaling Commission reported the figures last month. The captured whales included 181 females, 95% of which were pregnant. In one area of the hunt, more than half of the animals of both sexes were juveniles. The International Court of Justice temporarily banned Japan from whaling in



the Southern Ocean in 2014 (pictured, a whale captured by a Japanese whaling vessel), after deciding its hunts were not for scientific purposes as claimed. The nation launched a new whaling programme in 2015, called NEWREP-A.

## Forest fines

Brazil's environmental-protection agency has fined several agricultural companies for purchasing soya beans that were produced on illegally cleared land. Dubbed Operation Soy Sauce, the investigation by the Brazilian Institute of Environment and Renewable Natural Resources resulted in a total of 105.7 million reais (US\$28 million) in fines against five companies, including agricultural giants Cargill in Wayzata, Minnesota, and Bunge in White Plains, New York. The investigation identified illegal agricultural operations on 77 properties in

the Brazilian Cerrado, a stretch of savannah that borders the Amazon region. Authorities have seized more than 5,000 tonnes of soya beans.

## BUSINESS

## Fossil-power policy

US President Donald Trump has directed the Department of Energy (DOE) to take immediate steps to prevent utility companies from shutting down "fuel-secure" coal and nuclear power plants, the White House said on 1 June. Administration officials argue that impending retirements of such power plants for economic reasons are endangering national security, despite assurances from electricity-grid operators that there is no threat. In January, federal regulators rejected a DOE plan to subsidize the coal and nuclear industries, but the agency is now exploring legal strategies to compel grid operators to purchase electricity from troubled facilities, according to a leaked memo dated 29 May.

## POLITICS

## Minister resigns

Physicist Wu Maw-kuen resigned as Taiwan's education minister — who has responsibility for universities — on 29 May, after only 41 days in the position. In a statement to *Nature*, Wu

said that he was stepping down because Taiwan's opposition party had made "false accusations" against him that were interfering with the work of the education ministry. At a press conference in April, Hung Meng-kai of the nationalist Kuomintang (KMT) party alleged that Wu stole patented technology while he was president of Dong Hwa University between 2012 and 2016. In a statement to the media, Wu denied the allegation. On 25 May, KMT politician Ko Chih-en also alleged that when Wu was head of the National Science Council, he attended a conference in Hangzhou in 2005 without the government's permission, which is required of public servants. Wu told *Nature* that although he had attended the conference, he believes he had approval to do so.

## Italian government

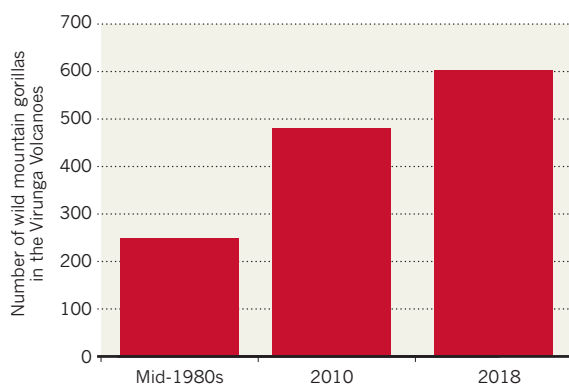
Two populist parties — the right-wing League party and the anti-establishment Five Star Movement — have formed a coalition government in Italy, ending months of political deadlock after an inconclusive election result in March. On 31 May, Italy's president, Sergio Mattarella, appointed Marco Bussetti, a former physical-education teacher, as education and science minister. Science and research were not major issues in the election campaign, but there are already clues to the government's leanings on some issues. The new health minister, physician Giulia Grillo, had campaigned to reverse a 2017 decree that made vaccinations compulsory for schoolchildren, but she announced on 4 June that the government would not immediately reverse it. The environment minister, Sergio Costa, a former general with the environmental arm of the military police, is known for his successful investigations of illegal toxic-waste dumping in the Naples region.

## TREND WATCH

The global population of critically endangered mountain gorillas (*Gorilla beringei beringei*) has hit 1,000. A 2-year survey, published on 31 May, found at least 604 individuals around the Virunga Volcanoes in east Africa. The only other place where mountain gorillas are known to survive is in Uganda's Bwindi Impenetrable National Park, where a 2012 census found about 400. The May finding represents a 26% increase over 6 years in total individuals, thanks to both population growth and improved survey methods.

## MOUNTAIN-GORILLA NUMBERS CLIMB

A census of endangered wild mountain gorillas in the Virunga Volcanoes in east Africa recorded a minimum of 604 individuals. Only one other population, of about 400 animals, is known to exist.



SOURCE: GO NATURE.COM/2M2X620

NATURE.COM

For daily news updates see:

www.nature.com/news

# NEWS IN FOCUS

**POLICY** Independent board reviews US environment agency decisions **p.15**

**FUNDING** European Research Council shows that high-risk research pays off **p.16**

**AWARDS** Million-dollar Kavli prize recognizes three CRISPR scientists **p.17**

**WORKFORCE** Diverse labs get a competitive edge in science **p.19**



ALAN BARNETT PHOTOGRAPHY



Neuroscientist Nikos Logothetis is a director at the Max Planck Institute for Biological Cybernetics, where he used to run a primate laboratory.

## ANIMAL EXPERIMENTS

# Scientists criticize handling of animal-welfare charges

*Some researchers are upset about how Germany's prestigious Max Planck Society has reacted to the indictment of a leading neuroscientist.*

BY ALISON ABBOTT

Scientists at one of Germany's leading neuroscience institutes say that their employer, the Max Planck Society (MPS), is failing in its responsibility to defend them against efforts by animal-rights activists to disrupt research.

The researchers outlined their criticisms in

two letters to MPS leadership seen by *Nature*, and in interviews.

Their concerns relate to the MPS's handling of a struggle between activists and Nikos Logothetis, a world-renowned neuroscientist who has been a director at the Max Planck Institute for Biological Cybernetics (MPI-Biocyb) in Tübingen since 1996. An expert in visual perception, Logothetis studies how

the brain makes sense of the world, and used to run a primate laboratory at MPI-Biocyb.

The MPS, which has an annual public budget of €1.8 billion (US\$2.1 billion), is Germany's most prestigious research organization, and runs 84 institutes and facilities.

The struggle began in September 2014, when a German television channel aired footage taken by an undercover animal-welfare ►

► activist who had infiltrated Logothetis's lab, purporting to show mistreatment of research monkeys.

Death threats and insults to Logothetis and his family followed — and in 2015, Logothetis decided to wind down his primate lab and replace it with a rodent facility.

Events came to a head on 20 February this year, when Logothetis was indicted for allegedly violating animal-protection laws. The charges, which Logothetis denies, stem from complaints made to police on the basis of the 2014 footage. A trial date has not yet been set.

## ANIMAL BAN

After the indictment, the MPS leadership removed Logothetis's overall responsibility for animal research at MPI-Biocyb, and banned him from conducting experiments with animals and from supervising others doing animal work. MPI-Biocyb scientists take issue with the MPS's decision to impose these sanctions on Logothetis before the case is considered by a court.

"We are very upset that the society is failing to uphold the principle of innocence before guilt is proven," says neuroscientist Hamid Noori, a junior group leader at MPI-Biocyb. "With this attitude, any activist can attack us freely, without consequence."

Noori is one of four MPI-Biocyb scientists who spoke to *Nature* about the situation. Their criticisms are echoed in the two letters: the first, sent in December, was signed by 54 scientists; the second, sent in February, was signed by 94, who make up the majority of those who work with animals at MPI-Biocyb, says Noori.

The February letter describes "an extremely distressful situation" that "has seriously compromised our working conditions".

MPS president Martin Stratmann says that the society, a publicly funded body, was justified in restricting Logothetis's responsibilities because it "must uphold public trust that animal research is carried out properly". Stratmann adds: "Any public perception that animals are being treated incorrectly will damage the image of animal research as a whole."

He says that he has met with staff to listen to their concerns, and that "there has been constant support for the Logothetis department and for Nikos Logothetis personally in the last years."

Stratmann also says that he has made

several public statements on the need for animal research in general and primate research in particular. But since 2014, Logothetis has claimed that the MPS's expressions of public support have not gone far enough.

The affair has caught the attention of the broader neuroscience community. "From the outside, it looks like the Max Planck Society is abdicating its duty to stand up for its scientists," says neuroscientist Bill Newsome at Stanford University in California, who co-chaired the US National Institutes of Health BRAIN Initiative when it was announced in 2013 by then-president Barack Obama. He says that the MPS "gave a negative message" by announcing sanctions before a verdict has been reached by a court.

The indictment follows contradictory judgments about Logothetis and his work. Immediately after the September 2014 documentary was broadcast, an external specialist appointed by the MPS leadership found no welfare violations at MPI-Biocyb. But two months later, the German Animal Welfare Federation, a non-

**"Our work was stopped and started, we were subjected to a lot of uncertainties."**

profit organization in Bonn, filed multiple complaints with police about animals at the institute.

In August last year, a local judge in Tübingen dismissed

all but one charge; for that charge, allegedly delaying euthanasia in three rhesus monkeys, the judge offered an out-of-court settlement, which Logothetis accepted. But in October, prosecutors in the state capital, Stuttgart, overturned the settlement decision. They pursued the delayed-euthanasia case against Logothetis and two other staff members, who have not been publicly named, leading to their indictment in February.

Logothetis says that the decisions about whether and when to kill the monkeys, which contracted infections after surgery, were appropriate and complied with the law. Veterinary staff attempted to treat the infections, he says, and two of the monkeys recovered. The third was humanely killed when staff decided that it was unlikely to recover.

## WORK PROBLEMS

Scientists at MPI-Biocyb say the MPS's handling of the saga has affected research. In October 2017, Stratmann cancelled a visit

of the institute's external scientific advisory board weeks before it was scheduled to happen, because of the ongoing case against Logothetis.

As a result, says chemist Goran Angelovski, a project leader at MPI-Biocyb who signed both letters to the society, scientists at MPI-Biocyb have not obtained the formal critiques of their work on which they rely to get promotions, or their next positions.

"The board's report would have given us an evaluation, and that's what we really need for our careers," he says. The short-notice cancellation also meant weeks of wasted work, according to those who signed the first letter.

Stratmann says that he takes these concerns very seriously, and that "preparations are currently ongoing to organize an evaluation". The preparation work for the cancelled visit can be used for this upcoming evaluation "to a large extent", he adds.

## RODENT PLANS HALTED

There have also been other disruptions, say scientists at MPI-Biocyb. In December last year, the MPS announced that Logothetis would lose his responsibility for animal research if he were indicted, and that the society was halting its plans to build the rodent-research facility as a result of the ongoing case. In January, the MPS announced that building work on at least a part of the facility would go ahead.

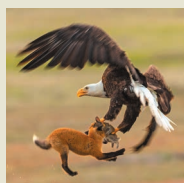
But these decisions sparked a period of disruption that has damaged productivity, says Henry Evrard, a neuroanatomist at MPI-Biocyb. "Our work was stopped and started, we were subjected to a lot of uncertainties," he says.

Stratmann says that Logothetis remains the scientific head of the institute's department of physiology of cognitive processes, and is still "able to plan, analyse and publish experiments". "The MPS has not taken away Nikos Logothetis's capacity to conduct research," he says. He emphasizes that the elements of the rodent facility that have been approved for construction are "very extensive".

Logothetis stresses his gratitude for the funding generosity of the MPS, "which offered me truly the whole universe". But he has appealed to a labour court for return of his full management responsibilities. "I need to clear my name," he says. ■



## IMAGES OF THE MONTH



Moons, memories and melees — May's best science images  
[go.nature.com/2xhkut1](http://go.nature.com/2xhkut1)

## MORE NEWS

- The researchers who study alien linguistics [go.nature.com/2kqrnc](http://go.nature.com/2kqrnc)
- Can UK science have its cake and eat it, too? [go.nature.com/2jvonky](http://go.nature.com/2jvonky)
- Dinosaur auction reflects rise in private fossil sales [go.nature.com/2hm5vcs](http://go.nature.com/2hm5vcs)

## NATURE PODCAST



Magnetic animal migration; cold enzymes; and mouse memory  
[nature.com/nature/podcast](http://nature.com/nature/podcast)

## ENVIRONMENT

# EPA data rule questioned

*Independent science board will review decisions by the US environment agency to repeal or change climate regulations and rules on the use of non-public data.*

BY JEFF TOLLEFSON

Science advisers to the US Environmental Protection Agency (EPA) voted on 31 May to review a series of controversial rules that the agency has proposed over the past eight months. These include a plan that would limit the types of scientific research that the EPA could use to justify environmental regulations, and proposals to strike down limits on greenhouse-gas emissions.

EPA administrator Scott Pruitt framed the data rule as part of a push for transparency — and against ‘secret science’ — when he released it on 24 April. The policy would prevent the EPA from relying on studies that include any data that have not been made public.

The decision by the EPA Science Advisory Board (SAB) to review the rule comes after earlier criticism by some of its members. In a 12 May memorandum, an SAB working group chastised the EPA for not submitting the proposal to the board for review.

“The working group is very much in favour of transparency,” said Alison Cullen, an environmental-health researcher at the University of Washington in Seattle, during the advisory board’s meeting. But on this particular proposal, there is a “very real lack of clarity” in how the rule would be applied, said Cullen, who chairs the working group.

The proposed transparency rule is modelled

on a similar bill that Republican lawmakers in the House of Representatives have pushed for years. The House passed the latest version of the legislation in 2017, but it died in the Senate.

Scientists and environmentalists have decried the EPA’s proposal, noting that many important epidemiological studies are based on public-health data that cannot legally be released owing to privacy concerns. As a result, critics say, such a rule would prevent the agency from considering some of the best health research, ultimately making it harder to create new environmental regulations.

Under previous presidents, the EPA has typically given the SAB advanced notice of regulatory actions, such as the release of a proposed rule, although that is not required by law. This week’s meeting was the first time that the full panel had considered the transparency rule. The EPA is not required to follow the advice of its advisory board, but failing to do so could bolster legal challenges against the agency.

The environment agency has yet to finalize the transparency rule: the deadline for public comments, originally scheduled to close on 30 May, has been extended to 16 August.

## EMISSIONS FIGHT

The science-advisory board also voted to assess the research underlying a series of proposed regulations to limit greenhouse-gas emissions from power plants, vehicles, and oil

and gas operations. That includes a review of the research behind Pruitt’s decision to repeal the Clean Power Plan. The plan sought to reduce carbon emissions from existing power plants and was former president Barack Obama’s signature climate-change policy. The advisers also intend to look over a decision made by the EPA in April to revoke emissions standards for vehicles manufactured between 2022 and 2025.

Separate emissions standards set by the state of California, and followed by a dozen other states, would remain in place; California

**“Any administration can reject our advice, but we are part of the record.”**

officials have warned that they will fight any attempt by Pruitt to revoke a waiver that allows the state to set its own regulations in this regard.

The EPA has yet to propose new standards to replace the Clean Power Plan or the Obama administration’s vehicle-emissions regulations.

The advisers did what they were supposed to do, said board member Steven Hamburg, chief scientist for the Environmental Defense Fund, an advocacy group based in New York City. “The SAB is a congressionally chartered organization,” he said. “Any administration can reject our advice, but we are part of the record.” ■

## ATMOSPHERIC SCIENCE

# Hurricanes around the world linger longer

*This means more rain and possibly more damage from storms.*

BY GIORGIA GUGLIELMI

Sluggish hurricanes have become increasingly common over the past 70 years, according to a new study. Storms that linger over a given area for longer periods, such as Hurricane Harvey, which stalled over eastern Texas for almost a week in August 2017, bring more rain and have greater potential to cause damage than ones that pass quickly. Scientists

are not sure why this is happening, but if the trend continues, future hurricanes could be even more disastrous.

The study, published this week in *Nature*<sup>1</sup>, is the first to analyse hurricane speeds globally. It finds that the speed at which tropical cyclones moved across the planet slowed by about 10% between 1949 and 2016: from more than 19 kilometres per hour on average in 1949, to about 17 kilometres per hour in 2016.

Over land, cyclones affecting regions along the western North Pacific slowed by 30%; over Australia and areas in or near the North Atlantic, they slowed by about 20%.

“That’s a big signal,” says study author James Kossin, a climate scientist at the US National Oceanic and Atmospheric Administration’s (NOAA) Cooperative Institute for Meteorological Satellite Studies in Madison, Wisconsin. Research suggested that atmospheric circulation patterns in the tropics might be slowing as a result of global warming, so Kossin set out to see whether hurricanes, which are carried along by these wind currents, have also slowed.

Because storms are becoming more sluggish, there’s more time for rain to fall. Kossin notes that a 10% reduction in hurricane speed corresponds to a 10% increase in the amount of rainfall over a given area. The effect could be magnified by a warming climate, because ►



Hurricane Harvey lingered over eastern Texas for days, flooding cities including Houston (pictured).

► recent global simulations estimate up to a 10% increase in rainfall per degree Celsius of warming.

Slower, more rain-heavy hurricanes would lead to more flooding events, says David Nolan, a hurricane scientist at the University of Miami in Florida. Stronger, more sustained winds are also more likely to damage buildings, he says.

The study results are interesting, says Tom Knutson, a research meteorologist at NOAA's Geophysical Fluid Dynamics Laboratory in Princeton, New Jersey. But researchers aren't sure what has caused the slowdown. Knutson says it's an open question whether human-driven climate change or natural variability is to blame. It's also unclear if the slowdown in atmospheric tropical circulation patterns

influences the speed at which hurricanes move across the globe. Knutson notes that his team's climate models, which simulate future Atlantic hurricanes, don't predict that storms will slow down — even when researchers tweak their models to slow those circulation patterns<sup>2</sup>.

The observed decrease in hurricane speed could be a result of unreliable data, says Kevin Trenberth, a climate scientist at the US National Center for Atmospheric Research in Boulder, Colorado. He notes that satellites have tracked storms across the globe only since the late 1960s, so data acquired before then might not be reliable and should be discounted.

But Kossin disagrees, saying that data on the speed of these storms are less sensitive to technological advances than data about their frequency and intensity. Moreover, he says, a study this year found that several past hurricanes would have been slower had they occurred in a warmer climate<sup>3</sup>. "That gives us more confidence that the slowing is there and is related to warming." ■

1. Kossin, J. *Nature* <https://doi.org/10.1038/s41586-018-0158-3> (2018).
2. Knutson, T. R. *et al. J. Clim.* **26**, 6591–6617 (2013).
3. Gutmann, E. D. *et al. J. Clim.* **31**, 3643–3657 (2018).

## RESEARCH GRANTS

# Europe's top funder shows high-risk research pays off

*European Research Council publishes third impact assessment of the projects it supports.*

BY INGA VESPER

A popular and unusual self-review carried out by Europe's most prestigious science funder is back. The annual assessment, now in its third year, found that nearly one in five projects supported by the European Research Council

(ERC) led to a scientific breakthrough.

The independent review, undertaken in 2017 and published on 31 May (see [go.nature.com/2jg2n3v](https://go.nature.com/2jg2n3v)), assessed 223 completed ERC projects that had ended by mid-2015. It concluded that 79% of them achieved a major scientific advance, 19% of which were considered fundamental breakthroughs. That proportion

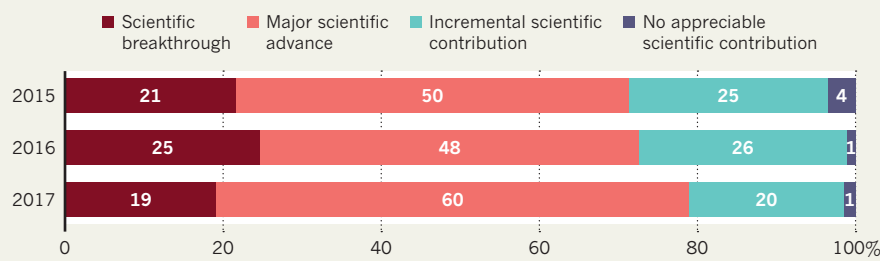
rose to 27% for ERC Advanced Grants, which are awarded to experienced researchers. Only 1% of the total were judged to have made no appreciable scientific contribution.

Established in 2007 to improve the quality of Europe's science, the ERC is the European Union's premier funder of blue-skies research and is part of Horizon 2020, the EU's main science-funding programme. It awards generous, multiyear grants in any discipline, and applications are judged solely on their quality. The council has undertaken annual reviews of the projects it funds since it ran a popular pilot assessment in 2015. That is a pioneering strategy among European funders, most of which evaluate success on a project-by-project basis, and it has been praised for taking a qualitative approach rather than relying, for instance, on bibliometrics.

The latest assessment was carried out by senior scientists convened by the ERC's Scientific Council. Each panel member was asked a series of questions about a randomly selected

## EUROPE'S TOP RESEARCH GRANTS

About one-fifth of projects funded by prestigious European Research Council grants make scientific breakthroughs, according to the council's qualitative self-assessments.



VILNIUS UNIV. (LEFT)/HALLBAUER&FORETTI  
FOTOGRAFIE (MIDDLE)/UC BERKELEY (RIGHT)

set of projects. This year, evaluators were also asked to focus on a project's risk to a greater extent than in previous years. (A spokesperson for the ERC said that the council is still refining the assessment's methodology.)

The 19% figure for scientific breakthroughs in the latest assessment is lower than in previous years: 21% and 25% of ERC projects assessed in the 2015 and 2016 exercises, respectively, were classed as such (see 'Europe's top research grants').

The reviewers concluded that most projects that made breakthroughs were high risk and high reward, and only 10%

**"The ERC has really pushed the expectation of taking more risks."**

of projects were considered low risk. "The ERC has really pushed the expectation of raising the boundaries of science and taking

more risks," says Jan Palmowski, secretary-general of the Guild of European Research-Intensive Universities, a lobby group in Brussels.

The assessment shows that risk-friendly funding is crucial for retaining talent in Europe, where research funders are generally risk-averse, says Martin Vechev, a computer scientist at the Swiss Federal Institute of Technology in Zurich who received an ERC grant aimed at early-career researchers in 2015, after spending time at computing firm IBM in the United States. The grant encouraged him to stay in Europe, and he says that the funding helped his team to develop a new subfield of artificial intelligence that focuses on machines that automatically write computer code.

The reviewers also said that more than 50% of projects had already made an economic and societal impact. In a speech earlier this year, ERC president Jean-Pierre Bourguignon said that council-funded research generated 29% of patents approved through EU funding in 2007–13, despite receiving less than 17% of the money.

#### FUNDING INCENTIVE

The review comes at a crucial time for EU research funding, say observers. This week, the European Commission is expected to release a detailed budget plan for the next instalment of its main funding programme, which will include the ERC's next pot of money. The programme, called Horizon Europe, will run from 2021 to 2027 and has a proposed budget of nearly €100 billion (US\$117 billion).

The latest review provides ammunition in the fight to raise the ERC's budget, says Palmowski. His organization is advocating for a doubling of the annual budget, which in 2017 was €1.8 billion; it started at €300 million in 2007. ■



Virginijus Siksnyš, Emmanuelle Charpentier and Jennifer Doudna won the 2018 Kavli nanoscience prize.

#### AWARDS

## Kavli prize recognizes scooped biochemist

*Virginijus Siksnyš shares award for CRISPR contributions.*

BY GIORGIA GUGLIELMI

CRISPR has hauled in yet another big science award, and this time the recognition includes a scientist whose contribution has sometimes been overlooked.

Two biochemists widely credited with co-inventing the gene-editing technology, Emmanuelle Charpentier and Jennifer Doudna, were named on 31 May as the winners of this year's Kavli Prize in Nanoscience. So was Virginijus Siksnyš, a Lithuanian biochemist whose independent work on CRISPR has thus far garnered much less mainstream attention — and Nobel-prize buzz — than that of Charpentier, Doudna and some other scientists.

Researchers working on the mechanism of hearing and on the formation of stars and planets also won Kavli prizes this year, in neuroscience and astrophysics, respectively.

The Kavli Foundation — established by the late Norwegian philanthropist Fred Kavli in Los Angeles, California — and the Norwegian Academy of Science and Letters in Oslo announced the three biennial prizes, each of which comes with US\$1 million to be split between the winners. First awarded in 2008, the prizes honour seminal research selected by three panels of experts from six global science societies and academies.

The nanoscience committee awarded the

prize to Charpentier at the Max Planck Institute for Infection Biology in Berlin, Doudna at the University of California, Berkeley (UC-Berkeley) and Siksnyš at Vilnius University in Lithuania "for the invention of CRISPR-Cas9, a precise nanotool for editing DNA, causing a revolution in biology, agriculture and medicine".

In 2012, a group led by Charpentier and Doudna<sup>1</sup>, and several months later one led by Siksnyš<sup>2</sup>, reported programming the CRISPR-Cas9 system to cut DNA at specific sites. Since then, award committees, the media and some in the scientific community have emphasized the roles of Doudna and Charpentier in developing the transformative gene-editing tool. In 2015, the pair shared the Breakthrough Prize in Life Sciences, worth \$3 million, for example.

But Siksnyš's work on CRISPR has occasionally been overlooked. The Kavli nanoscience committee recognized that the three researchers conducted "key pioneering work" in the development of CRISPR-based genome editing, says chairman Arne Brataas, a physicist at the Norwegian University of Science and Technology in Trondheim.

Siksnyš says he was "surprised" when a phone call from Oslo announced that he would share the Kavli prize with Doudna and Charpentier. "You don't expect such calls every day," he says. He adds that he is still disappointed it took so long to publish his results. *Cell* rejected his ▶

► paper in April 2012 without sending it out for peer review; Doudna and Charpentier submitted their work two months later to *Science*, which fast-tracked it for publication. The Kavli prize is “a good example that you can achieve your goals and get recognition”, Siksnys adds.

“This is a well-deserved prize for three individuals whose discovery made an enormous impact on modern biology,” says Rotem Sorek, a microbial geneticist at the Weizmann Institute of Science in Rehovot, Israel. Sorek is not surprised that Siksnys shared the prize. “In the field of CRISPR, he is well known as one of the pioneers of the technology.”

“It’s nice that the recognition is being spread around,” adds Dana Carroll, a biochemist at the University of Utah in Salt Lake City. But he notes that many others have also contributed to the development of CRISPR.

Synthetic biologist Feng Zhang, at the Broad Institute of MIT and Harvard in Cambridge, Massachusetts, has also shared in the accolades. His team was among the first to apply CRISPR gene-editing to mammalian cells, including mouse and human cells<sup>3</sup>. “There’s the tendency to pick the principal investigators,” says George Church, a geneticist at Harvard Medical School in Boston, Massachusetts, whose team also deployed CRISPR in human cells<sup>4</sup>. Young researchers such as the students and postdocs who turned CRISPR into a powerful

gene-editing tool tend to be ignored, he adds. But Church also says that more attention should be paid to other DNA-editing tools — for example, zinc-finger nucleases — some of which are already finding use in medicine and agriculture.

#### EARS AND STARS

The neuroscience award went to geneticist Christine Petit of the Pasteur Institute in Paris, and neuroscientists Robert Fettiplace at the University of Wisconsin–Madison and James Hudspeth at the Rockefeller University in New York City, “for their pioneering work on

**“This is a well-deserved prize for three individuals whose discovery made an enormous impact.”**

the molecular and neural mechanisms of hearing”. The researchers independently investigated the role of hair cells in the inner ear. These cells, which are covered

in microscopic hair-like projections, detect sound signals and transmit them to the brain<sup>5</sup>.

Ewine van Dishoeck, winner of the astrophysics category, works in astrochemistry at Leiden University in the Netherlands, where she has made her mark “elucidating the life cycle of interstellar clouds and the formation of stars and planets”, according to the prize citation.

Her work combines theoretical studies

with observations — especially, those made with infrared spectroscopy — and laboratory experiments to understand how compounds form in space, including the organic molecules that might have been the building blocks for life. She has also used radio telescopes to study planet formation around other stars. Van Dishoeck is the president-elect of the International Astronomical Union, and will lead celebrations next year as the union celebrates its 100th anniversary.

The laureates will receive their prizes in Oslo on 4 September. ■

1. Jinek, M. *et al. Science* **337**, 816–821 (2012).
2. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. *et al. Proc. Natl Acad. Sci. USA* **109**, E2579–E2586 (2012).
3. Cong, L. *et al. Science* **339**, 819–823 (2013).
4. Mali, P. *et al. Science* **339**, 823–826 (2013).
5. Fettiplace, R. *Compr. Physiol.* **7**, 1197–1227 (2017).

#### CORRECTION

The Editorial ‘Smelting point’ (*Nature* **557**, 280; 2018) misstated the amount of aluminium produced in 2017. It was more than 63 million tonnes, not 63,000.

And the Editorial ‘False testimony’ (*Nature* **557**, 612; 2018) gave the wrong numbers for the closed cases. The numbers of closed cases were 31 and 49, not 25 and 39.



# *The power of* diversity

BEING INCLUSIVE GIVES TEAMS A COMPETITIVE EDGE IN SCIENCE.  
IT ALSO HAPPENS TO BE THE RIGHT THING TO DO.

BY KENDALL POWELL

**C**hanel Phillips tried hard to keep her nerves in check last year as she prepared to present her research at an international meeting on drowning prevention. Phillips, a doctoral student at the University of Otago in Dunedin, New Zealand, was in Vancouver, Canada, discussing water-safety strategies that she had been developing with indigenous Maori communities back home. It struck her that she and a colleague were perhaps the only indigenous presenters at the meeting. ►

**Anne-Marie Jackson**  
is integrating Maori  
culture with science.

► She forged ahead, discussing unconventional work that blends Western and traditional Maori research methods to an audience of experts from around the world. Her presentation was warmly received. Researchers bombarded her with questions about working with indigenous groups — which often face rates of drowning that are higher than national averages. More important to Phillips, however, were the eager questions and comments coming from the Maori community members whom she worked with: she had live-streamed her talk so that they could listen and weigh in. “That was a big moment for me,” she says.

Back in Dunedin, Phillips is part of a group that supports young Maori researchers, who are under-represented in New Zealand science. Many countries are attempting to increase the diversity of their scientific workforces and to make the research they support better reflect the varied concerns of their populations. And the trend towards inclusion in science has also yielded benefits to the researchers who embrace it. A variety of studies have tracked different types of diversity — ethnic, gender, nationality and scientific discipline — and suggest that particularly diverse groups publish a higher number of papers and receive more citations per paper than average<sup>1,2</sup>. Diverse groups also seem to achieve better community participation when studying minority populations, and they often benefit from the different ideas and perspectives that the team members can bring.

*Nature* talked to three groups that have prioritized diversity in their ranks to ask about the benefits they have seen and the challenges and trade-offs they have to accept as part of the sometimes-difficult job of being inclusive (see page 149).

Although they differ greatly by geography and discipline, the groups share some key elements, including lab leaders who are directly engaged in the work, have high expectations and think that a diverse team produces the strongest research. These factors seem to be the secret sauce for these groups’ notable cohesion and robust research output.

For Anne-Marie Jackson, one of Phillips’s advisers, those benefits are secondary to the goal of creating science that will better serve the needs of indigenous communities. It’s science’s job to provide solutions for diverse communities, she says, so science must “reflect a diversity that comes from outside the ivory tower”.

## A CULTURE OF RESPECT

When Jackson, who studies health and physical education in Maori communities, joined the faculty at Otago in 2011, there was only one Maori graduate student in the programme. Jackson saw that as a serious problem. “The engine room of any research group, department or school is the graduate-student group,” she says. So, in 2013, Jackson teamed up with Hauiti Hākopa, a Maori cultural scientist, to create Te Koronga, a programme designed specifically for Maori graduate students.

The programme’s name is intended to reflect ‘a desire for higher learning’ and the group grounds students in Maori research methods alongside Western ones. Students learn lessons such as how to behave when asking questions in indigenous communities and how to interpret data from the ‘oral literature’ of traditional knowledge passed down in storytelling. Jackson usually has 10–15 honours, master’s and doctoral students in the Te Koronga group. At regular Monday meetings, they speak the Maori language *te reo Māori*, which many of the students are learning. The discussions help them to connect with traditional knowledge and legitimize both its place and theirs in academia, says Phillips.

Phillips’s work, for example, aims to strengthen Maori cultural connections to seafaring ancestors who reached New Zealand’s shores many generations ago. It includes reconnecting young people with traditional practices that elicit a strong sense of respect for the water. Phillips has found that that’s often a more successful strategy to introducing water safety than, say, first handing them a life jacket.

The group’s research centres on working with Maori communities to improve their well-being, not working ‘on’ them. This approach dovetails with recommendations made by Linda Tuhiwai Smith in her 1999 book, *Decolonizing Methodologies*. She argues that indigenous communities that have formerly been ‘the researched’ must now become ‘the researchers’ who set the agenda.

Such approaches promise to help fill a troubling gap in many areas of human research. For example, Esteban Burchard, Sam Oh at the University of California, San Francisco, and their collaborators showed in 2015 that fewer than 5% of federally funded studies on respiratory health in the United States represent people from minority groups<sup>3</sup>. “We’re using these studies to inform policy and drug development,” says Oh. “There’s a problem when we take this average data that is disproportionately white and then apply it to other populations.”

Jackson points out that there are other practical incentives for fostering cultural ties and a collaborative spirit with research community members.

Research that solves real-world problems earns a boost in the New Zealand system of performance-based research funding. And success is measured in part through statements from community members about how well programmes support their goals.

Phillips found, also, that live-streaming her international talk kept her accountable to her community. She is on track to be the first Te Koronga student to graduate with

a PhD, but all of those who have graduated with honours and master’s degrees have gone on to work in research, for government agencies or in Maori community programmes.

Jackson and Phillips both stress that Te Koronga is not about lowering standards for indigenous students to enter academic careers — a frequent argument from critics of such programmes. On the contrary, it demands academic excellence. Students need that to get in the front door, says Jackson, “but that is what our communities deserve, too — to have the best possible people working with them”.

## BORDERLESS RESEARCH

Materials scientist Mukhles Sowwan says he doesn’t actively try to recruit diverse, international researchers to his group — it just happens. In 2011, the Palestinian researcher moved from Al-Quds University in Jerusalem to the Okinawa Institute of Science and Technology Graduate University (OIST) in Japan, a fledgling start-up institute with a mandate to recruit across international borders.

“It’s a unique experiment and I wanted to be part of that,” Sowwan says, calling OIST “a perfect place to do research”.

OIST sits on the tropical island of Okinawa, which helps it to attract scientists from around the world. Its founding board, which included heavy-hitter Nobel laureates Sydney Brenner, Torsten Wiesel and Susumu Tonegawa, laid some ground rules to help OIST succeed.

At least 50% of researchers at every level — faculty members, postdocs, doctoral students — must be from outside Japan, and English is the official language for everything.

So far, the plan has paid off. OIST began admitting graduate students in 2012, and now has about 60 faculty members — recruited from every continent but Antarctica. It is currently ranked 119th out of 1,286 academic institutions in the Asia Pacific region and it is 8th among Japanese academic institutions according to its research output, as calculated by the *Nature* Index.

The founders encouraged diversity of all types — ethnic, gender, disciplinary and academic background. “There was a commitment to diversity with the expectation that it was, not just a nice thing to do, but necessary to the success of the whole thing,” says Robert Baughman, executive vice-president of OIST.

The data support that assertion. Several studies have found that a

*“I hope he will enter a world where neurodiversity is something people understand and celebrate.”*

SOURCE: REF. 4

greater mix of nationalities and ethnicities correlates with a lift in research output and with bigger impact through citations. In a 2013 analysis<sup>1</sup>, papers generated from the United States and the United Kingdom that had at least one international author garnered an “impact premium” — a relative bump in the average citation rate — over papers in which all authors were domestic.

A 2014 analysis<sup>2</sup> looked at 2.5 million US-based papers using surnames as a proxy (albeit a crude one) for ethnicity. Papers with 4 or 5 authors of different ethnicities had 5–10% more citations on average than papers from authors of all the same ethnicity.

Sowwan's group, which includes ten scientists from Africa, Asia, Australia, Europe and the United States, designs nanoparticles for environmental monitoring and biomedicine, blending materials science, physics, chemistry, nanodevice fabrication and computer simulation.

Sowwan says that the varied backgrounds of the lab members strengthen the group's output, and that he tends to see two tracks of scientists turn up in his lab. Many of the scientists recruited from world-class universities are ‘big picture’ thinkers, who, when given a particular research question, return with more excellent questions. The researchers who hail from countries with weaker scientific infrastructures, he says, generally excel at addressing every last detail of the original question.

“They complement each other, and, in the end, they will learn from each other how to always be looking in new directions and at the details,” says Sowwan.

In the past 2 years, Sowwan's group has published 14 papers and filed 4 US patents. And six former postdocs have since landed assistant-professor positions at universities around the world.

But the reasons for the connections between diversity and scientific reach and impact are not entirely clear. “It is easy to find correlations, but it is hard to prove causality,” says Wei Lee Woon, a computer scientist at Khalifa University of Science & Technology in Abu Dhabi.

It could be that it is not the diversity breeding the success, but rather that successful labs in top-tier institutions attract a diverse array of good candidates. To unpick the association, Woon and his Khalifa colleagues Talal Rahwan and Bedoor AlShebli probed the Microsoft Academic Graph, a database of scholarly publications. The trio analysed papers published between 1958 and 2009 by universities in the United States, United Kingdom, Canada and Australia<sup>4</sup>, assigning impact to about 1 million papers and their groups of co-authors on the basis of 5-year citation counts. The data scientists also used authors' surnames to gauge ethnicity, first names to infer gender and first publication date to assess academic age. They also accounted for authors' discipline and affiliation.

When they plotted those scores against the papers' citation counts, the ethnic diversity of the group correlated with impact more strongly than did any other category (see ‘Diversity's impact’). But did greater ethnic diversity actually drive citations up? To assess that, the researchers needed a “controlled experiment where we could fix everything else a priori and then ‘apply’ diversity to a set of papers”, says Rahwan.

The team split the papers into those with higher-than-average ethnic-diversity scores and those with lower-than-average scores. Crucially, the researchers ensured that the two sets of papers were similar in terms of publication year, number of authors per paper and the other four types of diversity. They matched 45,689 papers and say that they found that, all other factors being equal, the papers written by ethnically diverse groups were cited 11.2% more than were papers written by non-diverse groups.

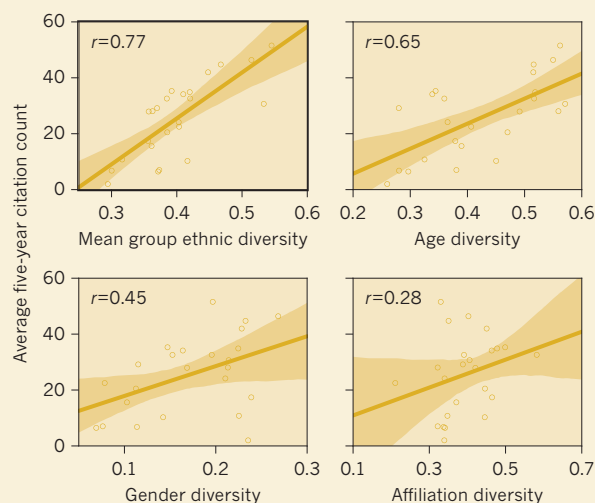
The team also tracked ‘individual’ diversity, using the ethnic diversity of each author's entire set of collaborators across all papers. Higher individual scores also gave higher citation counts, but group diversity was the bigger driving force behind a paper's impact<sup>4</sup>.

“Ethnic diversity is more about the now — who you are collaborating with right now,” on a paper, says AlShebli. Research labs or companies would be wise to hire researchers who complement the group's existing ethnic mix, adds Woon.

The mechanism behind the bump is still elusive. “We've had so many discussions about this,” says Woon. It could still be down to high-quality leaders attracting the best people from around the world. Or papers with more international authors could be shared with a wider, more diverse ►

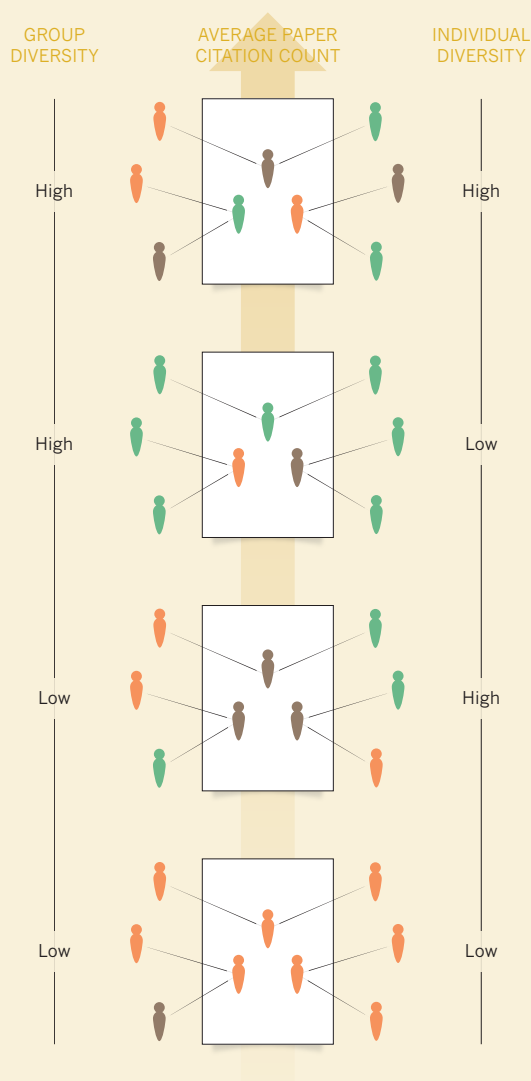
## DIVERSITY'S IMPACT

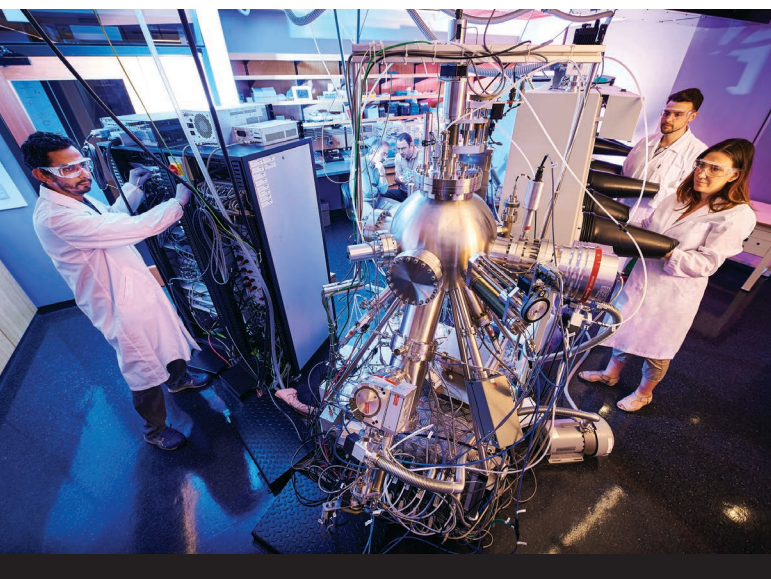
Ethnic diversity correlates more strongly ( $r$ ) with citation counts than do diversity in age, gender or affiliation, according to an analysis of more than 1 million papers in 24 academic subfields (circles).



### GROUP VERSUS INDIVIDUAL DIVERSITY

Diversity within the list of authors on a paper (group diversity) has a stronger effect on citation count than does individual diversity, the measure of diversity in a researcher's network of collaborators.





► network, increasing citations.

Woon's team will probe the mechanism of wider networks next. But, Woon says, he would be a little disappointed if that turned out to be the answer to diversity's positive influence. "We all hope it's due to the different cultures cross-fertilizing ideas."

### EMBRACING THE SPECTRUM

In his doctorate work at Vanderbilt University in Nashville, Tennessee, Dave Caudel says, he drew on his unique way of processing the world. As someone with autism spectrum disorder (ASD), he found a welcoming home in the lab of astrophysicist Keivan Stassun, who has been an unflinching advocate of diversity in science. Stassun thinks that diverse perspectives increase the frequency of 'aha' moments.

Caudel finds no reason to argue with that. "It's hard to point to something that was valuable because he was black or she was a Hispanic woman," he says. "But it's no accident that Keivan has published in *Nature* five times. That's him and his group harnessing all that capability."

Although Caudel had published as an undergraduate researcher, he bombed in the standardized tests needed for entry into graduate school — scoring in the bottom 4%. He found his way to Stassun's group through the Fisk–Vanderbilt Bridge Program, an initiative to help under-represented groups succeed in science. The bridge programme "opened a door where there was none", Caudel says.

Stassun's open-door approach springs from his personal background, growing up Mexican American with a single mother and starting his education, he says, from the "bottom of the socio-economic ladder".

Now senior associate dean for graduate education in the College of Arts and Sciences at Vanderbilt, he welcomes young scientists from groups that are under-represented in astrophysics, including women and ethnic minorities, people from low-income families and first-generation university students. Currently, he's mentoring 19 graduate students and postdocs from very different backgrounds.

"I'm intentionally bringing in people from diverse backgrounds and with diverse ways of thinking because it forces us all out of easy assumptions," he says.

Caudel is part of Stassun's efforts to embrace neurodiversity — which describes people whose brains process information differently owing to conditions such as ASD, attention deficit hyperactivity disorder or dyslexia. Once again, the drive to be inclusive stems from Stassun's own experience — he has a son with ASD who dreams of becoming an engineer. "I hope he will enter a world where neurodiversity is something people understand and celebrate."

Caudel relishes working in a lab where he is celebrated for his differences and "given the freedom to work in the way that works best for my brain and how I'm wired". And when Caudel misses a social cue or

Muhkles Sowwan's team includes scientists from five continents.

fails to interpret a facial expression, the group is very accepting.

Caudel's way of processing information has brought a huge advantage to the work, says Stassun. Caudel says that he can spend hours, days, weeks or even months thinking obsessively about a problem, "chipping away at the series of thoughts". But he also has times when he gets overloaded with too much sensory input and has to take a day off.

Ironically, Stassun points out, a welcoming, vibrant lab that is socially buzzing and chatty can be difficult for some people with ASD to navigate. So he has come up with different ways for people to participate — such as replacing face-to-face attendance at big group meetings with instant messaging.

The group remains highly productive. It published 47 papers in 2017. Stassun calls it a "well-oiled machine with all the oars pulling in the same direction at the same time".

He points to a 2013 discovery showing that the size and evolutionary stage of a star could be determined by a clever measurement of its flickering starlight<sup>5</sup>. "A multiracial, multigendered group of researchers put our heads together and made this very surprising discovery," he says. "It required a group of very different kinds of people looking at the same data from different perspectives."

### CHALLENGES

Tensions are inevitable when running diverse research teams, where lab mates bump up against each other's cultural differences, language barriers and varying beliefs on what constitutes 'personal space'. In Stassun's group, the level of diversity forces everyone to "take an extra beat" and think twice before speaking in vernacular, short-hand or culturally coded language. Stassun says his biggest hurdle is teaching his lab members to be better, more precise communicators.

"This isn't just me being nice" to guard against misstatements or offensive remarks, he says. "This lab requires precise communication to be at its best."

Some lab heads 'force' their groups to interact socially and scientifically — sometimes by literally rubbing elbows, sitting in close proximity. The Stassun lab has weekly, somewhat mandatory, social gatherings that include food and drink. And Sowwan's team never misses a chance to celebrate a birthday or to share cultural celebrations such as Diwali.

Lab members say that the together time helps to dissolve cultural biases and misunderstandings. The top-down leadership of their lab heads sets the expectation that even if people are uncomfortable at times with diversity, it's worth the fresh perspectives that it brings.

Sowwan sees the frequent, mini culture clashes that happen at OIST as an opportunity to attack the problem of biases from the bottom up. OIST researchers share a camaraderie because "we are all from outside", says Marta Haro, an electrochemist in Sowwan's lab. Her colleague Zakaria Ziadi, an electrical engineer, says that the atmosphere breeds tolerance. "Working here, you see people becoming more flexible, adaptable and more accepting."

The cohesion of each of these groups embodies something referred to as critical mass — a phenomenon in which under-represented groups experience less stereotyping and higher inclusivity when their numbers reach a certain percentage, typically 15–30%, of the total work group<sup>6</sup>.

More investigators should follow the lead of inclusive labs because the pay-off in scientific impact is real and substantial, says Rahwan. "People need to go beyond their comfort zones, because there is a measurable benefit." ■ **SEE EDITORIAL P.5 AND CAREERS P.149**

**Kendall Powell** is a freelance writer based in Lafayette, Colorado.

1. Adams, J. *Nature* **497**, 557–560 (2013).
2. Freeman, R. B. & Huang, W. *Nature* **513**, 305 (2014).
3. Burchard, E. G., Oh, S. S., Foreman, M. G. & Celedón, J. C. *Am. J. Respir. Crit. Care Med.* **191**, 514–521 (2015).
4. AlShebli, B. K., Rahwan, T. & Woon, W. L. Preprint at <https://arxiv.org/abs/1803.02282> (2018).
5. Bastien, F. A., Stassun, K. G., Basri, G. & Pepper, J. *Nature* **500**, 427–430 (2013).
6. Nielsen, M. W. et al. *Proc. Natl Acad. Sci. USA* **114**, 1740–1742 (2017).

# COMMENT

**GEOSCIENCE** Did the Anthropocene begin in the 1600s? **p.26**



**GENETICS** Dark start of heredity science traced to asylum statistics **p.28**

**ARCTIC** Early-career climate-science collaborations transcend tensions **p.30**

**EARTH OBSERVATION** AI experts needed to map road-building boom **p.30**

ROMAN UCHYTEL



An artist's impression of the extinct woolly rhinoceros (*Coelodonta antiquitatis*).

## Big data little help in megafauna mysteries

Too many meta-analyses of extinctions of giant kangaroos or huge sloths use data that are poor or poorly understood, warn **Gilbert J. Price** and colleagues.

**I**n March, the last male northern white rhinoceros died. The sub-species joins a long list of large land animals that have gone extinct over the past 100,000 years.

The reason for the demise of the northern white rhinoceros (*Ceratotherium simum cottoni*) is undisputed: poaching and land disturbance by people. By contrast, who or what caused the extinctions of mammoths, enormous ground sloths and other Quaternary megafauna remains one of the most contested topics in the historical sciences.

Was the culprit early humans who dispersed from Africa more than 75,000 years ago? Or was it climate change? The latest way to try to settle the debate involves meta-analyses. These attempt to link the timing of extinctions to shifts in the climate, or to evidence of the first appearance of humans in a particular region. Over the past five years, the number of meta-analyses has greatly increased (see 'In fashion'). Many have been published in high-impact journals, and they are starting to shape the debate.

Understanding why some groups succumbed while others survived could provide insights into how modern-day species might — or might not — survive climatic and environmental changes, and into the resilience of natural ecosystems to increasing anthropogenic impact.

But in our view, the 'big-data' approach cannot, at this point, get us closer to an answer. There simply aren't enough good-quality data. An understanding of what drove the extinctions requires detailed analysis ►

► on a species-by-species basis. This means investing effort into finding more fossil specimens and verifying the ages of those that have already been discovered using improved dating methods. It also means relating the timing of species' existence and disappearance to detailed local environmental, climatic and archaeological records.

### HUMAN LINK

For a typical meta-analysis, researchers mine the literature for dates associated with now-extinct megafauna, as well as for estimates of when humans arrived at a particular region (on the basis of archaeological and other data). In some cases, they then combine these records with global-scale palaeoclimate data, such as those obtained from ice cores collected from the Arctic. By mapping correlations between events, investigators try to identify the dominant factor driving species losses.

Over the past two decades, most of the meta-analyses that merge continental or global-scale data sets have pointed the finger at modern humans. In fact, some researchers contend that the results are so clear that there is no need for further debate<sup>1</sup>.

For any meta-analysis, however, the reliability of results is largely governed by the 'GIGO' principle: garbage in, garbage out. In our view, most of these analyses depend on questionable data, making the results hard to interpret at best. Six key problems undermine many of the studies conducted so far.

### Outdated geochronological information.

Models frequently use data from studies that have been superseded. For instance, during the 1980s, radiocarbon dating of species such as the Eurasian woolly rhinoceros (*Coelodonta antiquitatis*) suggested that it survived well into the Holocene — perhaps until as recently as 3,600 years ago<sup>2</sup>. But refinements in dating methods have shown that the rhinos had actually disappeared by about 14,000 years ago<sup>3</sup>. Some of the most recent big-data studies still use erroneous early dates for the rhino<sup>4</sup> and other species<sup>5</sup>.

**“With good data, models could provide crucial insights about large-scale changes.”**

**Contested dates.** In other cases, the dates associated with certain species are still in question. For instance, researchers first estimated the age of the elephant-like *Stegodon trigonocephalus* not by dating the fossils themselves, but by dating fossils from other animals collected from deposits more than 100 kilometres away<sup>6</sup>. Other investigators have flagged problems associated with using inferred ages<sup>7</sup>, yet these continue to be fed into meta-analyses<sup>8</sup>.

In some cases, ages are assigned to species



Artist's impression of the extinct land crocodile, a giant kangaroo and a giant wombat-like marsupial.

that have never even been dated, directly or indirectly. A 2016 study<sup>4</sup>, for instance, listed Australian animals such as the land crocodile *Quinkana* and the giant wombat *Ramsayia* among the megafauna thought to have existed in the past 100,000 years. The fossils of these species have never been dated<sup>9</sup>. (More than 25 of Australia's megafaunal species, or around 30%, have never been dated, simply because no one has done the work.)

**Insufficient data.** Some meta-analyses take the last appearance of a species in the fossil record to be the time when the animal went extinct<sup>7</sup>. In the rare cases where hundreds of samples have been found, for instance for mammoths and mastodons, a species' disappearance from the fossil record could well signal its demise. Yet where only a few specimens exist, the last appearance in the fossil record might have little bearing on the timing of the extinction.

A step in the right direction are probabilistic models of extinction times. These incorporate a degree of error associated with the age of specimens, based in part on the quality of the methods used to date them. Again, the robustness of the results depends on the quality of the data fed in. At this point, very few of the species that went extinct over the past 100,000 years are associated with reliable dates<sup>10</sup>. (In our view, the cave lion (*Panthera spelaea*), woolly rhino and woolly mammoth (*Mammuthus primigenius*) are among the handful of species for which sufficient data exist to enable a modelling approach.)

**Problematic proxies.** In the absence of fossil bones, some researchers have used proxy data to test megafaunal extinction hypotheses. For instance, the coprophilous fungus *Sporormiella* is a common component of

the pollen and spore fossil record. Because it occurs on animal dung, an abundance of it in a sediment core is often taken to indicate high numbers of big herbivores. Some investigators assume that a decrease in the appearance of the fungus over time and its eventual disappearance from the fossil record signal the extinction of megafauna<sup>11</sup>.

Yet *Sporormiella* lives on the excrement of a vast range of both big and small animals, including mammals and birds, herbivores and even some carnivores<sup>12</sup>. Its abundance is also affected by factors such as climate and water flow. Thus, on its own, levels of *Sporormiella* in a pollen core can't provide information about which species were present at any one time, or in what numbers.

**Insufficient scrutiny.** Lastly, long lists of extinct species (frequently just names and numbers in supplementary materials) often do not receive the necessary level of scrutiny. This has led to some unfortunate errors. The authors of at least two studies<sup>4,13</sup> have argued, for instance, that *Homo sapiens* caused the demise of giant marsupials such as *Euryzygomia dunense* and *Euowenia grata*. These were extinct for millions of years before *Homo sapiens* even appeared; they are known only from the Pliocene, the period 5.3 million to 2.6 million years ago. Another paper<sup>14</sup> suggested that the genus *Macropus* went extinct in Australia some 40,000 years ago. In fact, *Macropus* is alive and kicking: it includes Australia's extant kangaroos.

**Arbitrary definition.** Megafauna are commonly defined as Quaternary terrestrial vertebrates with a mass of at least 44 kilograms — roughly 100 pounds. This is a nice, round cut-off, but it is essentially arbitrary. Also, in some cases, 'megafauna'

are not so mega. For instance, they could include extinct terrestrial vertebrates that are larger than their extant cousins but that weigh considerably less than 44 kilograms. An extinct relative of the modern-day Australian echidna — *Megalibgwilia ramsayi* — is considered to be megafauna, even though it weighed only around 15 kilograms when it existed (until at least around 100,000 years ago).

In other words, megafauna are highly biologically and ecologically diverse, with several species separated from each other by hundreds of millions of years of evolution. Researchers should not therefore expect them to have responded in the same way to changes in their environments — whether driven by humans or by climate.

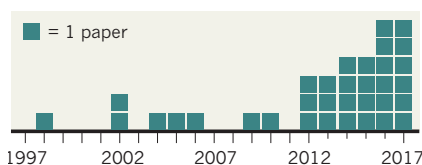
### A BETTER WAY

We think that as long as data from the fossil record remain scant, an understanding of what drove the extinctions of large animals over the past 100,000 years requires detailed analysis on a species-by-species basis. This means trying to find new fossils and verifying the estimated age of specimens previously found — for instance, through repeated sampling, or by using improved techniques to date museum specimens.

It also means taking into account all the local palaeoenvironmental information that is available to develop a detailed understanding of the palaeoecology of each species and its ecosystem. To reconstruct the diet of an animal, researchers can use stable isotope analyses of tooth enamel. Pollen cores can indicate the local vegetation at the time. The geochemistry of certain formations nearby, such as stalagmites, might give clues about the local climate. Changes in the nature of the sediment laid down in a nearby creek bed,

### IN FASHION

Meta-analyses are increasingly being used to study the drivers of past extinctions of big animals.



\*Studies included are those that have modelled large data sets of published ages.

or in the deposition of sand dunes, might hint at local landscape changes. And so on. Broad global palaeo-temperature records are likely to be a crude guide to climatic and environmental changes at local scales.

For each species, investigators should also strive to develop a clearer understanding of the human populations that lived alongside, and the nature of their interactions. This could be obtained by analysing DNA samples extracted from ancient human remains, for instance, or by studying middens, ancient dumps for domestic waste.

For example a study published earlier this year combined new dating approaches with chemical analyses of the bones of the cave bear (*Ursus spelaeus*), to show that its herbivorous diet had remained unchanged up until its last appearance in Europe, some 23,500 years ago<sup>15</sup>. Moreover, cut marks on its bones have revealed that some of these animals were hunted by humans. And researchers have linked the morphology of the extinct eastern African antelope *Damaliscus hypsodon* to the open, dry grasslands it inhabited, to track the demise of both<sup>16</sup>.

Megafaunal fossils can now be dated with much greater efficiency and precision — including those of animals that existed several hundreds of thousands

of years ago. This is thanks to various advances, such as combined U-series and electron-spin resonance dating. Other emerging techniques, such as the extraction and analysis of ancient DNA, can shed light on changes to the population size of now-extinct species. Several studies have used such approaches to demonstrate that populations of taxa, from giant Irish elk (*Megaloceros giganteus*)<sup>17</sup> to the Beringian steppe bison (*Bison priscus*)<sup>18</sup>, plummeted many thousands of years before their ultimate extinction, apparently because of deteriorating local climates and habitat changes.

Some might counter that we're averse to change and are simply finding another reason to be alarmed about the demise of the field sciences in a digital world<sup>19</sup>. But our argument is not with modelling per se. With good data, models could provide crucial insights into large-scale changes and the broad nature of the interactions between humans and other big animals as humans dispersed from Africa. More data, of better quality, can be obtained only through fieldwork and rigorous analysis of fossil materials. ■

**Gilbert J. Price** is a senior lecturer in palaeontology at The University of Queensland, Australia. **Julien Louys** is an Australian Research Council Future Fellow at Griffith University, Queensland, Australia. **J. Tyler Faith** is a curator of archaeology and an assistant professor of anthropology at the University of Utah in Salt Lake City. **Eline Lorenzen** is a curator and associate professor at the Natural History Museum of Denmark, København. **Michael C. Westaway** is a senior research fellow at Griffith University. e-mail: g.price1@uq.edu.au

1. Flannery, T. Q. *Essay* **48**, 1–80 (2012).
2. Turvey, S. T., Tong, H., Stuart, A. J. & Lister, A. M. *Quat. Sci. Rev.* **76**, 156–166 (2013).
3. Stuart, A. J. & Lister, A. M. *Quat. Sci. Rev.* **51**, 1–17 (2012).
4. Bartlett, L. J. et al. *Ecography* **39**, 152–161 (2016).
5. Wan, X. & Zhang, Z. *Proc. R. Soc. B* **284**, 20162438 (2017).
6. van den Bergh, G. D., de Vos, J. & Sondaar, P. Y. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **171**, 385–408 (2001).
7. Westaway, M. C. & Groves, C. P. *Archaeol. Oceania* **44**, 84–95 (2009).
8. Sandom, C., Faurby, S., Sandel, B. & Svenning, J.-C. *Proc. R. Soc. B* **281**, 20133254 (2014).
9. Wroe, S. et al. *Proc. Natl Acad. Sci.* **110**, 8777–8781 (2013).
10. Stuart, A. J. *Geol. J.* **50**, 338–363 (2015).
11. Gill, J. L., Williams, J. W., Jackson, S. T., Lininger, K. B. & Robinson, G. S. *Science* **326**, 1100–1103 (2009).
12. Ahmed, S. I. & Cain, R. F. *Can. J. Bot.* **50**, 419–477 (1972).
13. Johnson, C. N. *Proc. R. Soc. B: Biol. Sci.* **269**, 2221–2227 (2002).
14. Salter, F. et al. *Nature Commun.* **7**, 10511 (2016).
15. Terlato, G. et al. *Histor. Biol.* <https://doi.org/10.1080/008912963.2018.1448395> (2018).
16. Faith, J. T. et al. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **361–362**, 84–93 (2012).
17. Cooper, A. et al. *Science* **349**, 602–606 (2015).
18. Shapiro, B. et al. *Science* **306**, 1561–1565 (2004).
19. Ríos-Saldaña, C. A., Delibes-Mateos, M. & Ferreira, C. C. *Glob. Ecol. Conserv.* **14**, e00389 (2018).



The giant wombat-like *Euryzygoma* went extinct long before *Homo sapiens* even evolved.



## CLIMATE SCIENCE

# Seeking the Anthropocene

Wolfgang Lucht examines a book linking the contested epoch to early globalization.

In 1864, palaeontologist Édouard Lartet made a stunning discovery. At La Madeleine in southern France, workers had unearthed a few fragments of mammoth ivory engraved with a vividly detailed depiction of the animal itself. Here, finally, was proof that humans had seen the mammoth. The artefact also implied something more disturbing: that Earth's climate is not as stable as had been thought, and that species coexisting with humans can become extinct.

In *The Human Planet*, geographer Simon Lewis and geologist Mark Maslin provide a compelling narrative, stretching from the

emergence of hominins from Earth's long history some 3 million years ago, to our position today, as a species with planetary reach. Explaining the many ways in which we are now profoundly altering Earth, from polar melt to deforestation, they provide convincing evidence that we should indeed dub our new epoch the Anthropocene. Trying to understand our era, they observe, means parsing “a heady mix of science, politics, philosophy and religion linked to our deepest fears and utopian visions”.

The geological division of time into epochs, as they note, “is a human construct,

created to help make sense of the world we find ourselves in”. The Anthropocene is largely uncontested as a phenomenon. But formalizing it — its definition and when it began — is hotly debated, a conversation Lewis and Maslin have been involved in for some time (S. L. Lewis and M. A. Maslin *Nature* **519**, 171–180; 2015). The debate has raged since Earth-system scientist and Nobel laureate Paul Crutzen established the idea of the Anthropocene almost two decades ago, following on from much older considerations of a human-dominated geological era. Also party to the protracted skirmish are

A Svalbard reindeer on drifting Arctic ice.



around 12,000 years ago; as Maslin and Lewis point out, geologically it is just one more interglacial period in a series reaching back 2.6 million years. However, it does coincide with the rise of human civilization, and the environmental impacts that eventually followed. Given that we are now an emerging meta-civilization with planetary consequences, should the Holocene have been called the Anthropocene? And if so, when, precisely, did the Anthropocene begin — and how might that beginning be visible in future geological deposits?

Lewis and Maslin aim to clear the mists by establishing criteria for the Anthropocene's status as a geological epoch. They first outline four crucial revolutions in the evolution of civilization that could be signature events

**"The changes we are imposing on the biosphere are the most long-lasting and troubling impact humans will have on Earth."**

for our progressive dominance of the planet. Two of these are the rise of agriculture from 11,000 to 5,000 BC, and of industrialization from the eighteenth century to today. These both greatly increased access to energy and resources; yet they also locked societies into consumption-related dependencies and feedbacks that are not readily broken. And the rise of capitalism amplified trade and the flow of information, causing revolutions that boosted first ecological homogenization, then socio-cultural globalization. Together, these processes led to the vast environmental shifts we see today.

The authors point to a hallmark incident of one of these revolutions, the first contact between Europe and the Americas, as the event that produced a suitable marker for the start of the Anthropocene. By 1610, a small but pronounced dip in atmospheric carbon dioxide concentration had occurred, detectable in Antarctic ice cores. This, they argue, was triggered by a chain of events over the previous century, starting with the collapse of indigenous American populations in the face of introduced diseases and violent suppression. Much land then went uncultivated; forests regrew and more carbon was sequestered.

The 1610 geoscientific signal is significant in context because it also marks a turning point in human-driven homogenization of the global ecology. Before European explorers reached the Americas, ecosystems had been separated by the Atlantic and Pacific oceans. Suddenly they were linked, leading to extensive exchanges of species, with evolutionarily significant consequences that will be reflected in the geological layers.

From a scientific standpoint, Lewis and Maslin paint their picture with an often amazingly broad brush. The jury is

still out among carbon-cycle experts (see M. Rubino *Nature Geosci.* **9**, 691–694; 2016) on whether the 1610 dip really is a signal of the environmental impact of European contact. It occurred during the Little Ice Age, a phase of planetary cooling that could have led to similar shifts in the atmosphere's carbon balance.

Equally, not all readers may be convinced by the authors' idea that males among early humans had lower testosterone levels than Neanderthals, and that this helped to make *Homo sapiens* society more cooperative, and allowed them to prevail over Neanderthals.

Readers wishing to dig deeper might turn to Clive Finlayson's *The Humans Who Went Extinct* (2009). This portrays *Homo sapiens* more convincingly, as a species that got lucky by being well adapted to ecological niches such as the steppes (which were extensive during the last ice age), just as Neanderthal populations fragmented. Those wishing to consider the palaeoanthropological foundations of the Anthropocene might wish to consult Clive Gamble's important *Origins and Revolutions* (2007). That examines how material culture has ever expressed human identity; there have been no 'revolutions' in culture, but rather a continuous process. Or they might access Steven Mithen's 1996 *The Prehistory of the Mind*, on how a fusion of social and technological mental capacities led to fluid intelligence. The unsurpassed 'Earth System Analysis — The Scope of the Challenge' by Hans Joachim Schellnhuber (a chapter in the 1998 *Earth System Analysis*, which he edited) offers a deeply co-evolutionary view of sustainability. And Tim Lenton and Andrew Watson's *Revolutions that Made the Earth* (2011) reveals Earth's history as a staggering interplay of planetary biogeochemistry and evolutionary transitions.

Nonetheless, *The Human Planet* is immensely readable and introduces important concepts. I agree with the authors' insistence that it would be wrong to base the Anthropocene on a largely climatic definition that depends on whether or not Earth's near-future state still qualifies as an interglacial. The changes we are imposing on the biosphere are the most long-lasting and troubling impact humans will have on Earth.

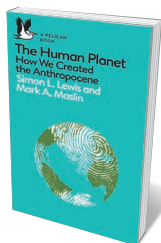
The Anthropocene debate is profoundly about how we see ourselves, and what we might do next. If we believe that scientific knowledge is universal, it is uniquely suited to informing a story applicable to all of humanity. We might do well to become '*Homo geosapiens*' by drawing conclusions from that story. ■

**Wolfgang Lucht** is an Earth-system scientist at the Potsdam Institute for Climate Impact Research and at Humboldt University in Berlin.  
e-mail: wolfgang.lucht@pik-potsdam.de

geologists representing the Subcommittee on Quaternary Stratigraphy of the International Union of Geological Sciences.

Indeed, designating the Anthropocene scientifically is a formidable task. Any marker for the beginning of major human impacts on the planet needs to be globally synchronized in the geological record. It also has to describe a process that casts a long shadow into the future history of Earth — enough to produce rock strata that mark a turn in Earth's planetary trajectory.

The designation of the Holocene complicates the debate. This universally agreed current epoch started



**The Human Planet: How We Created the Anthropocene**  
SIMON L. LEWIS & MARK A. MASLIN  
Pelican (2018)



The scene in Bethlem asylum, London, in William Hogarth's 1735 *A Rake's Progress*.

## GENETICS

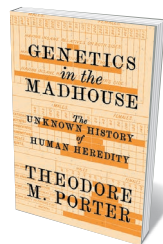
# The debt of genetics to mental illness

David Dobbs lauds a history tracing heredity science to statistics hoarded in asylums 250 years ago.

Who founded genetics? The line-up usually numbers four. William Bateson and Wilhelm Johannsen coined the terms genetics and gene, respectively, at the turn of the twentieth century. In 1910, Thomas Hunt Morgan began showing genetics at work in fruit flies (see E. Callaway *Nature* 516, 169; 2014). The runaway favourite is generally Gregor Mendel, who, in the mid-nineteenth century, crossbred pea plants to discover the basic rules of heredity.

Bosh, says historian Theodore Porter. These works are not the fount of genetics, but a rill distracting us from a much darker source: the statistical study of heredity in asylums for people with mental illnesses in late-eighteenth- and early-nineteenth-century Britain, wider Europe and the United States. There, “amid the moans, stench, and unruly despair of mostly hidden places where data were recorded, combined, and grouped into tables and graphs”, the first systematic theory of mental illness as hereditary emerged.

For more than 200 years, Porter argues in *Genetics in the Madhouse*, we have failed to recognize this wellspring of genetics — and thus to fully understand this discipline, which still dominates many individual and societal



**Genetics in the Madhouse: The Unknown History of Human Heredity**  
THEODORE M. PORTER  
Princeton University Press (2018)

Thurnam, who made the York Retreat in England a “model of statistical recording”. Better-known figures, such as statistician Karl Pearson and zoologist Charles Davenport — both ardent eugenicists — come later.

Inevitably, study methods changed over time. The early handwritten correlation tables and pedigrees of patients gave way to more elaborate statistical tools, genetic theory and today’s massive gene-association studies. Yet the imperatives and assumptions of that

responses to mental illness and diversity.

The study of heredity emerged, Porter argues, not as a science drawn to statistics, but as an international endeavour to mine data for associations to explain mental illness. Few recall most of the discipline’s early leaders, such as French psychiatrist, or ‘alienist’, Étienne Esquirol; and physician John

scattered early network of alienists remain intact in the big-data genomics of precision medicine, asserts Porter. And whether applied in 1820 or 2018, this approach too readily elevates biology over culture and statistics over context — and opens the door to eugenics.

As Porter notes, alert readers might ask how a force so crucial in the birth of genetics remained hidden. His answer distills to three arguments. First, after the Second World War, geneticists took pains to distance themselves from asylum science and eugenics, and historians left this largely unquestioned. Second, the system’s influence is partly obscured by its cruelty and neglect; we must look past those to see its determination to use statistics to identify people as ‘defective’. Third, the asylum network was easy to overlook because it was loose and decentralized.

It started with fine intentions. Many asylum founders of the late eighteenth and early nineteenth centuries hoped to cure people of mental illness through a humane, psychosocial “moral therapy”. These included the York Retreat’s founder, William Tuke, and Esquirol and his mentor, Phillipe Pinel.

These asylums and their records soon received transformative scrutiny. In 1788, King George III of Britain, who since his coronation had sometimes displayed symptoms suggesting psychosis, had an extreme episode. Understanding mental illness became a national-security issue. The alienists’ assessment, bolstered by physician William Black’s “original, useful, and authentick” statistics from London’s Bethlem asylum, gave the government leverage to replace the king with a regent — his son, later King George IV.

This much-publicized process spurred a decades-long growth in asylums run by the “numerical method”, and the use of national censuses to measure what seemed an epidemic of ‘insanity’. At the time, this baggy term encompassed a range of behaviours deemed extreme. Similar developments elsewhere helped spread this methodology across most of the developed Western world.

From London and Paris to Schussenried, Germany, and Worcester, Massachusetts, asylums grew and new ones sprouted. Knitting them together was an active system of correspondence, travel, conferences and publications such as the *American Journal of Insanity* and the *Allgemeine Zeitschrift für Psychiatrie*.

At first, asylums claimed absurdly high ‘cure’ rates. Reports of 50% were routine. The Connecticut Retreat claimed 91.6% four years in a row. By the mid-nineteenth century, however, asylum directors realized that they could simply say, as some big-data psychiatric geneticists do now, that although a cure seems distant, statistical patterns discovered in ever-larger study populations will one day reveal a cause — and a cure will follow. Funders bought it. Asylum science grew apace.

Eventually, having eliminated religious fervour, heartbreak, financial strain and

masturbation as causes for mental illness, alienists fixed on the only pattern left: patients' pedigrees. Heredity was "the one great cause ... the cause of causes", as French surgeon Ulysse Trélat proclaimed in 1856.

Thus asylum scientists unwittingly laid a path to disaster. For if mental illness boiled down to heredity, the final cure — if you insisted on imposing one — became both obvious and unspeakable.

Porter's chapters, with titles smacking of gothic Victorian novels, trace the long walk to corruption. 'Narratives of mad despair accumulate as information' gives way to 'German doctors organize data to turn the tables on degeneration, a foretaste of horror. The final chapter, 'Psychiatric geneticists create colossal databases, some with horrifying purposes, 1920–1939', sees eugenics deployed en masse. After the 1927 Supreme Court decision *Buck v. Bell*, US programmes forced sterilization on tens of thousands of people deemed mentally deficient. The Nazis built on that example in the 1930s by sterilizing some 400,000 Germans labelled hereditarily 'defective'. In 1940, they launched their wider genocidal programme by gathering more than 10,000 people from asylums all over southern Germany and gassing them at Grafeneck Castle.

The story of the era, Porter insists, is not one "of isolated failings by a few bad scientists". Every genetic insight along the way was sucked into the stream. Many geneticists and alienists had invested too heavily to stop. Others had the task brought to them. It was not by chance that the Holocaust found its first victims in asylums, which also housed the rosters, records and rationale that doomed them.

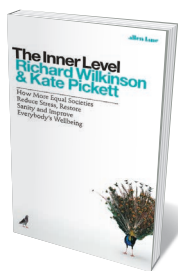
This matters for many reasons, according to Porter, the most immediate being the elemental links between this history and contemporary study of heredity. As Porter exposes strand after strand of connection, he draws sobering parallels between the motives, methods, obsessions and promises of bygone asylum directors, and those of the enormous human-genomics institutes that now enjoy unprecedented funding and power.

To Porter, these connections are roots, and today's genomics industry the tree. "Sold with a promise to find the genes for talents, diseases, and every kind of personal characteristic", he writes, genetics has returned to "the tradition of amassing, ordering, and depicting data of biological inheritance" that started more than two centuries ago, in squalor.

Some will reject this idea ferociously. But I suspect this bold, dauntingly well-documented book will prove difficult to dismiss. ■

**David Dobbs**, author of *My Mother's Lover*, writes on science, culture, music and sport for publications including *The New York Times*, *National Geographic*, *WIRED* and *The Atlantic*. His work can be found at [neuronculture.com](http://neuronculture.com).  
e-mail: [david.a.dobbs@gmail.com](mailto:david.a.dobbs@gmail.com)

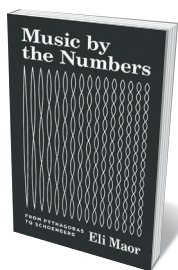
## Books in brief



### The Inner Level

Richard Wilkinson and Kate Pickett ALLEN LANE (2018)

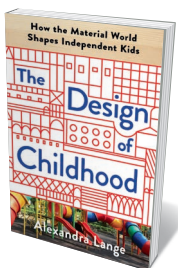
In *The Spirit Level* (2009), epidemiologists Richard Wilkinson and Kate Pickett probed the powerful correlation between a society's inequality and indices of well-being such as social mobility. Here, they narrow the focus to individuals. Drawing on wide-ranging research, they examine how inequity unsticks communities, leading to status anxiety, isolation, depression and rampant consumerism. They lay out pragmatic means of democratizing labour and dismantling class distinctions. And they put forth a salient point: that ability is generally a product, rather than a determinant, of social position.



### Music by the Numbers: from Pythagoras to Schoenberg

Eli Maor PRINCETON UNIVERSITY PRESS (2018)

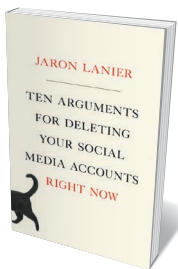
From precise notation to rhythmic patterns, music and mathematics often chime. In this intriguing study, maths historian Eli Maor traces those echoes, along with the trajectories of the "scientists, inventors, composers, and occasional eccentrics" behind them. We encounter the musical 'firsts' of classical philosopher Pythagoras; composer Arnold Schoenberg, whose "relativistic" music might have been influenced by the theories of Albert Einstein; the German musicians who in 2001 launched a 639-year performance of John Cage's composition 'As Slow as Possible'; and scores more.



### The Design of Childhood

Alexandra Lange BLOOMSBURY (2018)

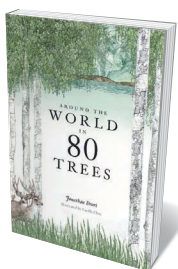
Millions of children are in digital overdrive, risking limited interaction with the material world (see B. Kiser *et al. Nature* **523**, 286–289; 2015). Alexandra Lange reminds us why that is an issue. Her captivating design history begins with construction toys such as Lego, Tubation and Zoob, and moves through home, school and playground as they morph to accommodate children's needs and inspire their creativity ever more fluidly and beautifully. She shows, too, how in mixed urban spaces, child-centred elements such as play areas and mental-mapping landmarks are often elbowed out.



### Ten Arguments for Deleting Your Social Media Accounts Right Now

Jaron Lanier HENRY HOLT (2018)

Fiercely unequivocal and utterly timely, Jaron Lanier's manifesto urges those still in thrall to social media to bin their accounts — now. The virtual-reality pioneer (see A. Faisal *Nature* **551**, 298–299; 2017) lays out ten rationales, starting baldly with "You are losing your free will". His argument, as an insider's insider, is that these "social modification empires" undermine truth, destroy empathy, promote unhappiness and make a joke of politics through constant surveillance and manipulation. As he puts it, it's better to be a cat, autonomous and in charge, than a subservient dog — or lab rat.



### Around the World in 80 Trees

Jonathan Drori and Lucille Clerc LAWRENCE KING (2018)

This tome, gorgeously illustrated by Lucille Clerc, pays homage to the tree as a scientific subject, a cultural mainstay and an exemplar of biological majesty. Educator Jonathan Drori has isolated 80 species for his global survey, each wreathed in intriguing tales. Blossoms of the long-lived lime (*Tilia x europaea*), for instance, exude the bee-befuddling sugar mannose, and seedpods of the Costa Rican sandbox (*Hura crepitans*) explode with the sound of a pistol shot, ejecting their load at up to 240 kilometres an hour. From upas to coco de mer, an arboreal odyssey. [Barbara Kiser](#)

# Correspondence

## Arctic collaboration transcends tensions

Of all world regions, the Arctic is the most sensitive to climate change and drives feedbacks that amplify the effects of global warming around the planet. Understanding the Arctic relies on developing a better knowledge of the hugely expansive Russian Arctic regions, which offer unique opportunities to study landscape systems across large latitudinal gradients, linked by major river networks.

However, these regions have been something of a blind spot for the international community of Arctic scientists. This is due to access difficulties and to research findings going unrecognized because of language barriers. Happily, at a time of mounting political tension between Russia and the United Kingdom, early-career Arctic scientists from both countries are working together.

Following workshops held in March at Lomonosov Moscow State University and at the British Antarctic Survey in Cambridge, UK, a group of these researchers are now collaborating to remove logistical hurdles and to combine complementary resources and expertise. The workshops were organized by the UK Natural Environment Research Council's Arctic Office, the UK Polar Network, the Association of Polar Early Career Scientists in Russia and the UK Science and Innovation Network.

Immediate challenges include pooling knowledge that is scattered among publications in English or in Russian. Imminent outcomes include the organization of a conference in the Russian Arctic, a database of funding sources, and guidelines for working in the area (see [go.nature.com/2jvdtntk](http://go.nature.com/2jvdtntk)).

This successful collaboration demonstrates how science diplomacy can transcend the hostility of government politics. Such cooperation among early-career researchers now should advance scientific and social

progress over the decades to come.

**Sammie Buzzard** *University College London, UK.*

**Joseph Cook** *University of Sheffield, UK.*

**Alexey Maslakov** *Lomonosov Moscow State University, Russia*  
[s.buzzard@ucl.ac.uk](mailto:s.buzzard@ucl.ac.uk)

## Road mapping needs AI experts

As road building expands globally, an automated system for detecting and mapping roads in near-real time is urgently needed to plan land use and conservation management. Machine-learning or artificial-intelligence (AI) specialists must help to meet this formidable challenge.

Current road data are grossly inadequate (see W. F. Laurance *et al. Nature* **513**, 229–232; 2014), and most mapping techniques rely on visual interpretation by humans. Even the community-led OpenStreetMap initiative — which aims to maintain accurate maps of roads and many other built features — is patchy and suffers from systematic biases among nations, regions and biomes ([www.openstreetmap.org](http://www.openstreetmap.org)). The freely available, high-resolution radar data sets being collected globally in all weather conditions under the European Union's Copernicus Earth-observation programme are an important advance.

What we sorely need now is a road-detection algorithm that can discriminate between paved and unpaved roads. Crucially, it would need to operate consistently under varying topographical and environmental conditions, and be able to distinguish roads from other linear components such as low walls, irrigation ditches and natural features.

A road-detection algorithm would be instrumental in the discovery of illegal roads that are imperilling the world's most vulnerable ecosystems and species. Authorities will then stand a fighting chance

of tackling this huge problem effectively.

**William F. Laurance** *James Cook University, Cairns, Australia.*  
[bill.laurance@jcu.edu.au](mailto:bill.laurance@jcu.edu.au)

## Circular economy creates new jobs

Governments are anticipating that people will be displaced from factory and service jobs as intelligent systems are increasingly deployed. Smart environmental enterprises could offer a more sustainable approach than solutions such as universal pay, and provide employment.

In a circular economy (see [www.nature.com/theirculareconomy](http://www.nature.com/theirculareconomy)), commercial enterprises that reverse environmental damage, for example, are needed to deliver value in a new guise. This can take the form of a surcharge for certifying improved environmental conditions — for instance, as a result of companies reusing materials from landfill, or cleaning up the environment after manufacturing processes.

As with other start-ups, government support would be essential. Once in motion, these companies could follow their own competitive paths. Such ventures would also encourage sustainability efforts among conventional manufacturers.

We urgently need to apply such business concepts to sustainability. Landfill space is running out and recyclable materials are piling up through a lack of capacity for handling and processing. The proposed enterprises would incorporate design and business-development functions geared to launching products and services based on waste and recyclable materials.

Some business areas and jobs would have to be created because they do not yet exist. For example, we need an effective process for removing microplastics from soil, water and air.

**Andrew Kusiak** *The University of Iowa, Iowa City, USA.*  
[andrew-kusiak@uiowa.edu](mailto:andrew-kusiak@uiowa.edu)

## Happy 250th to a Colombian great

This year is the 250th anniversary of the birth of the Colombian scientist, inventor, naturalist and astronomer, Francisco José de Caldas, who directed South America's first astronomical observatory, in Bogotá (then in New Granada). In 1816, a Spanish general had Caldas shot in the back for supporting independence of his country from Spain, claiming that "Spain does not need savants". The eminent scientist's legacy proved otherwise.

His pioneering contributions include the invention of the hypsometer for measuring altitude, based on his observation that the boiling temperature of distilled water is proportional to atmospheric pressure. He put this device to good use as a participant in the Botanical Expedition of New Granada (present-day Colombia, Ecuador, Panama and Venezuela), discovering that organisms adapt along altitudinal gradients in tropical ecosystems.

A passion for meteorology led Caldas to publish a paper in 1808 in which he explored the influence of the physical environment on human behaviour (F. J. D. Caldas *Semanario Nuevo Reino Granada* **22**, 200–207; 1808). His weather observations also enabled others to pin down the exact date of a volcanic eruption in the tropics in 1808, which had been suspected from sulfur isotopes in ice cores but was not recorded by eyewitnesses at the time (see A. Guevara-Murua *et al. Clim. Past* **10**, 1707–1722; 2014).

In the end, astronomy was his undoing. His Newtonian views on the Universe were considered heretical in the Spanish colonies at that time. And Caldas used the observatory as a cover for his revolutionary activities.

**César Marín** *Austral University of Chile, Valdivia, Chile.*  
[cesar.marin@postgrado.uach.cl](mailto:cesar.marin@postgrado.uach.cl)

## MAMMALIAN EVOLUTION

# A 3D view of early mammals

The unexpected discovery of a nearly complete skull from the Early Cretaceous epoch that has been preserved in three dimensions provides profound insights into the evolution and biogeography of early mammals. [SEE LETTER P.108](#)

SIMONE HOFFMANN & DAVID W. KRAUSE

The fossils that advance our understanding of evolutionary history often come from parts of the world that have not been well studied by palaeontologists. Occasionally, however, game-changing fossils are discovered in heavily surveyed regions by targeting poorly sampled rock layers in those areas. On page 108, Huttenlocker *et al.*<sup>1</sup> present one such discovery from the Cedar Mountain Formation in North America — an Early Cretaceous rock formation in Utah dated to between 139 million and 124 million years ago. The authors describe a complete, 3D fossil skull from a previously unknown genus and species, which they name *Cifelliodon wahkarmoosuch*.

Palaeontological fieldwork in western North America has led to the discovery of a greater number of Cretaceous mammal species and their more primitive relatives (collectively called mammaliaforms) here than in any other region in the world — more than 150 species have been found<sup>2</sup>. They are represented by tens of thousands of specimens, the vast majority of which are isolated teeth from the Late Cretaceous epoch (100 million to 66 million years ago). A few of the specimens are lower jaws, and many fewer are skulls or skeletons. The extreme rarity of 3D skulls makes Huttenlocker and colleagues' discovery momentous.

Aside from its 3D preservation (Fig. 1), the skull is remarkable in other ways. It is about 7 centimetres long, indicating that *Cifelliodon* was about the size of a medium hare. This would have been unusual in an Early Cretaceous world dominated by shrew- and mouse-sized mammals. Among its North American contemporaries, only one known species is larger — the carnivorous *Gobiconodon ostromi*<sup>2</sup>. In addition, the downturned face and relatively shallow snout make the skull unusual among early mammaliaforms.

Huttenlocker *et al.* used an imaging technique called micro-computed tomography ( $\mu$ CT) to reveal a wealth of anatomical detail about the skull. For example, they found that *Cifelliodon* had a small brain with large olfactory bulbs. This combination is commonly seen in early mammaliaforms<sup>3</sup>, and is indicative of the keen sense of smell that is



**Figure 1 | A fossil from the dawn of mammals.** Huttenlocker *et al.*<sup>1</sup> report the discovery of a nearly complete skull from the Cedar Mountain Formation in North America, dated to between 139 million and 124 million years ago. They named the species *Cifelliodon wahkarmoosuch*, and suggest that it belongs to a group of animals called haramiyidans. The skull is shown in side view with the snout pointing left. Scale bar, 10 millimetres. (Image adapted from Fig. 1 of ref. 1.)

one of the hallmarks of mammals today.

Using the same technique, the authors also demonstrated that the bones of the occipital region at the back of the skull (in particular, the tabular bones) remained unfused in *Cifelliodon*. Unfused tabulars are a rare trait in mammaliaforms — the tabulars are generally fused with the occipital bones in adult mammals, but remain separate in other vertebrates<sup>4</sup>. But, as the authors point out, the supposition that these bones are fused in early mammaliaforms could be biased by the rarity of well-preserved skulls, or by the techniques used to analyse them. The few mammaliaforms that have been found to have unfused tabular bones, as in *Cifelliodon*, are known from 3D skulls or have been scanned using  $\mu$ CT<sup>1,5</sup>, which enables the detection of cranial structures that might be invisible on the skull's surface.

The completeness of the *Cifelliodon* skull allowed Huttenlocker *et al.* to assess the early branches of the mammaliaform family tree more robustly than has previously been possible. This phylogenetic analysis places *Cifelliodon* in the extinct Haramiyida — a group that has elicited controversy because its position in the evolutionary tree has implications for the timing of the origin of mammals.

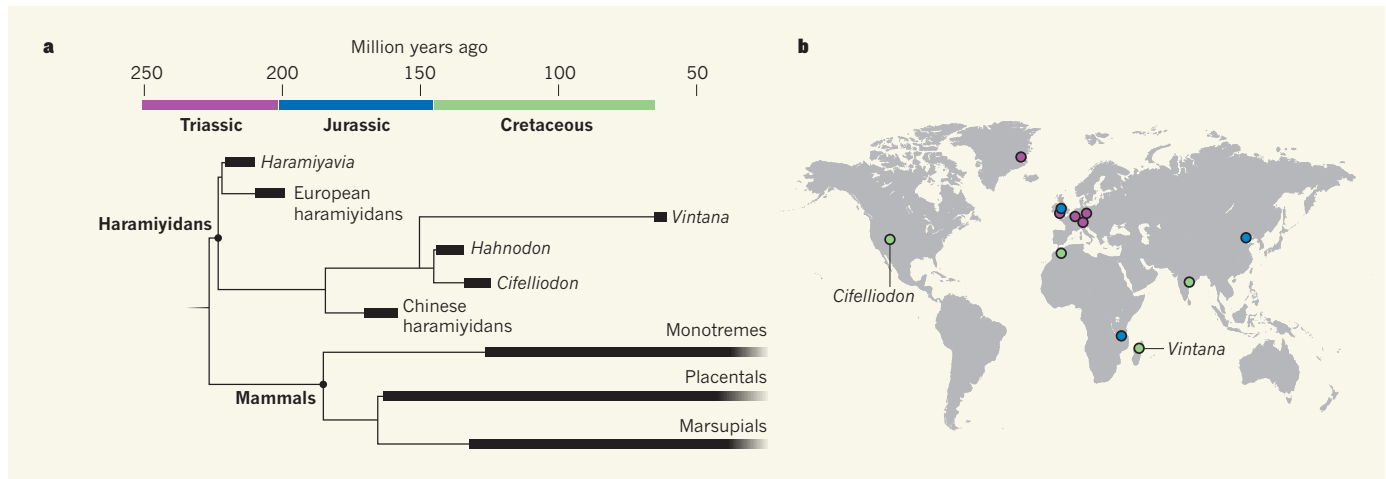
Until five years ago, the fossil record of haramiyidans consisted mainly of isolated teeth from a few European sites from the Late Triassic (237 million to 201 million years ago) and

Early Jurassic (201 million to 174 million years ago), and a lower jaw and a few postcranial bones from a Late Triassic site in Greenland<sup>2</sup>. In 2013, two almost-complete haramiyidan skeletons from the Middle Jurassic (174 million to 164 million years ago) were discovered in China, sparking controversy about mammalian relationships<sup>6,7</sup>. The two research groups involved came to very different conclusions: one placed Haramiyida in Mammalia<sup>6</sup>, the other outside it<sup>7</sup>. This difference is not trivial: it results in vastly different temporal estimates for the origin of mammals.

Placing Haramiyida in Mammalia pushes the origin of the latter group back to the Late Triassic (215 million years ago). Such a date would imply that several mammalian clades, including the lineage that led to placental and marsupial mammals, have earlier origins than was thought. This earlier date also implies that there were long intervals of time in which these early lineages were present but for which fossils have not yet been found.

By contrast, placing Haramiyida outside Mammalia suggests an origin in the Early Jurassic (about 185 million years ago). This implies a relatively explosive diversification of early mammals.

Several further haramiyidans from the Jurassic period have since been discovered in China, but the controversy has only intensified<sup>5,8–10</sup>. Particularly problematic is the fact



**Figure 2 | Re-evaluating the evolution and biogeography of haramiyidans.** **a**, Huttenlocker *et al.*<sup>1</sup> analysed relationships between the early branches of the family tree for mammals and their more primitive relatives. The resulting evolutionary tree indicates that haramiyidans are not mammals, contrary to some previous evidence<sup>5,6,8,9</sup>. The analysis also places the Cretaceous genus *Vintana* in Haramiyida for the first time. **b**, Cretaceous haramiyidans (indicated by green circles) have previously been found in northern Africa and possibly

India. The authors' analysis expands the Cretaceous range of haramiyidans to Madagascar (*Vintana*) and North America (*Cifelliodon*). Combined with the fact that other fossils of haramiyidans from the Triassic (purple) have been found in Europe and Greenland, and that haramiyidans from the Jurassic (blue) have been found in Europe, China and Tanzania, this work implies a much broader temporal and geographical distribution of haramiyidans than had previously been hypothesized.

that, although the Chinese haramiyidans are represented by complete skeletons, the specimens are essentially 2D. Most of the skulls are little more than flattened outlines, which limits their usefulness for informing mammalian relationships.

*Cifelliodon* is one of the first skulls preserved in three dimensions from the haramiyidan lineage. As such, it is a crucial piece of the evolutionary puzzle. Huttenlocker and colleagues' phylogeny puts Haramiyida (and so *Cifelliodon*) outside Mammalia (Fig. 2a). Thus, their work favours a model in which early mammals diversified rapidly during the Jurassic.

Finally, Huttenlocker *et al.* provide evidence that *Cifelliodon* is closely related to Cretaceous species from northern Africa (*Hahnodon taqueti*) and Madagascar (*Vintana sertichi*), the latter of which had not previously been assigned to Haramiyida. This implies a much broader temporal and geographical distribution for Haramiyida than has been assumed (Fig. 2b), indicating the need to reassess the biogeographical history of the group. The authors conclude that haramiyidans had a global distribution during the Jurassic–Cretaceous transition, and that land bridges aiding vertebrate dispersal existed long after the fragmentation of the supercontinent Pangaea — much later than previously recognized. An alternative hypothesis that is perhaps more consistent with current palaeogeographical models<sup>11</sup> is that haramiyidans, like many vertebrate groups, had a Pangaean distribution in the Jurassic period and evolved in isolation thereafter, as landmasses separated during the Cretaceous period. The best way to test these competing hypotheses is with the discovery of more well-preserved fossils, like this exquisite skull of *Cifelliodon*. ■

**Simone Hoffmann** is in the Department of Anatomy, New York Institute of Technology, College of Osteopathic Medicine, Old Westbury, New York 11568, USA. **David W. Krause** is in the Department of Earth Sciences, Denver Museum of Nature & Science, Denver, Colorado 80205, USA. e-mails: shoffm04@nyit.edu; david.krause@dmns.org

1. Huttenlocker, A. K., Grossnickle, D. M., Kirkland, J. I., Schultz, J. A. & Luo, Z.-X. *Nature* **558**, 108–112 (2018).
2. Kielan-Jaworowska, Z., Cifelli, R. L. & Luo, Z.-X. *Mammals from the Age of Dinosaurs: Origins, Evolution, and Structure* (Columbia Univ. Press, 2004).

3. Rowe, T. B., Macrini, T. E. & Luo, Z.-X. *Science* **332**, 955–957 (2011).
4. Koyabu, D., Maier, W. & Sánchez-Villagra, M. R. *Proc. Natl Acad. Sci. USA* **109**, 14075–14080 (2012).
5. Krause, D. W. *et al. Nature* **515**, 512–517 (2014).
6. Zheng, X., Bi, S., Wang, X. & Meng, J. *Nature* **500**, 199–202 (2013).
7. Zhou, C.-F., Wu, S., Martin, T. & Luo, Z.-X. *Nature* **500**, 163–167 (2013).
8. Bi, S., Wang, Y., Guan, J., Sheng, X. & Meng, J. *Nature* **514**, 579–584 (2014).
9. Han, G. *et al. Nature* **551**, 451–456 (2017).
10. Luo, Z.-X. *et al. Nature* **548**, 326–329 (2017).
11. Torsvik, T. H. & Cocks, R. M. *Earth History and Palaeogeography* (Cambridge Univ. Press, 2017).

This article was published online on 23 May 2018.

#### ASTRONOMY

## A fresh approach to stellar benchmarking

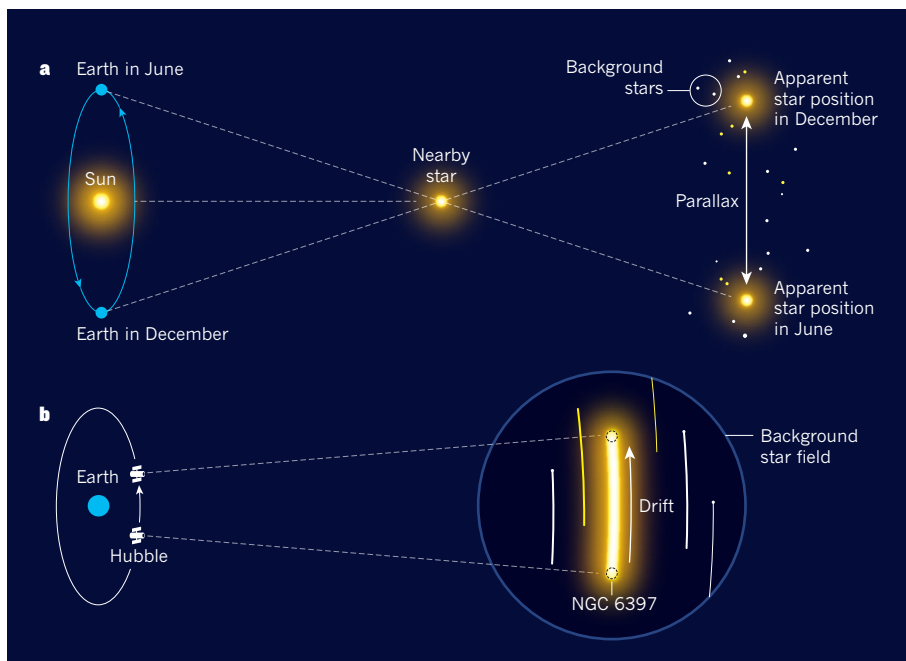
**An avalanche of data is about to revolutionize astronomy, but the options for validating those data have been limited. High-precision measurements from the Hubble Space Telescope enable a much-needed alternative option.**

**RACHAEL BEATON**

**T**ry this experiment: extend your thumb at arm's length and close one eye at a time. Your thumb will seem to 'jump' between two positions as you switch the eye that is closed. That jump is known as parallax. If you measure the jump as well as the distance between your eyes, you can use trigonometry to calculate the distance to your thumb. Astronomers use parallax, on a much greater

scale, to measure distances to astronomical objects. Writing in *The Astrophysical Journal Letters*, Brown *et al.*<sup>1</sup> report that they have achieved this for the nearby star cluster NGC 6397, using the Hubble Space Telescope. Their method will provide a crucial means of validating the wealth of parallax data released this year from the European Space Agency's Gaia mission<sup>2</sup>.

It is a challenge to find a topic in astronomy that does not rely on the astronomical distance



**Figure 1 | Using parallax to measure astronomical distances.** Objects viewed along two lines of sight have different apparent positions relative to their background, and the distance between those positions is known as the parallax. **a**, Distances to a nearby star are determined by measuring the parallax between two images of the star that were taken six months apart. Because the distance between the two observation positions is known to be the diameter of Earth's orbit around the Sun, the distance to the star can be calculated using trigonometry. **b**, Brown *et al.*<sup>1</sup> have measured the distance to the nearby star cluster NGC 6397. A camera on the Hubble Space Telescope took two long exposures six months apart, so that the cluster was visible as a 'streak' that results from the telescope's orbital motion around Earth (the apparent drift of the cluster has been exaggerated, for clarity). Each position along the streak provides a different measurement of the position of each star in the cluster, thereby allowing the apparent positions of the stars to be measured more precisely than from 'snapshot' images. These measurements enabled the distance to the cluster to be determined.

scale — a collection of methods applied in series to determine distances that are too large to be measured directly. Distances are used as conversion factors for deriving the physical quantities of celestial objects from observations and, therefore, they are essential for constructing models of the Universe. The foundations of the astronomical distance scale are trigonometric parallaxes for individual stars. These parallaxes enable us to calibrate the physical properties of those stars, which can then be used to infer properties of ever more distant stars, star clusters and galaxies. On the largest distance scales, they can even be used to calculate the size of the Universe.

A stellar parallax was first measured in 1838 by the astronomer Friedrich Bessel<sup>3</sup>. Many more have been recorded since. Until earlier this year, about 2 million reliable measurements<sup>4</sup> had been made. This sounds like an impressive number, but effectively spanned only the astronomical cul-de-sac in which the Sun resides. That number increased to roughly 1 billion following the release in April of data from Gaia<sup>2</sup>, which surveyed a region well beyond the Solar System, almost halfway across the Galaxy. Until the publication of Brown and colleagues' data, only one technique — very-long-baseline interferometry<sup>5</sup> — was capable of measuring parallax

directly on such distance scales. This was a concern because astronomers worldwide are poised to use the Gaia data in their research, and so it would be desirable to have more than one direct method for measuring stellar parallaxes to help validate the Gaia data.

The basic experimental set-up for measuring stellar parallax is identical to that described for observing your thumb. First, two images of the same astronomical object are taken with an interval of six months (Fig. 1a). This ensures that they are captured at positions separated by the diameter of Earth's orbit around the Sun, in the same way that observing your thumb from each eye provides two viewpoints that are separated by a known distance. Second, the apparent displacement of the target star is determined. This involves measuring the position of the star in each image with high precision, and then measuring that position in relation to a set of reference objects (stars or galaxies) from the same image. Both tasks are conceptually simple yet tricky to achieve in practice. Brown *et al.* address them in interesting ways.

To measure the position of their target star cluster, Brown and colleagues took the two images with a camera on the Hubble Space Telescope using a long exposure, so that the cluster's stars 'drift' across the images as a result of the telescope's orbital motion around Earth

(Fig. 1b). This technique, known as spatial scanning<sup>6,7</sup>, produces images of the target as a 'streak'. Every position along the streak provides a different measurement of the position of each star in the cluster.

The images of NGC 6397 taken by Brown *et al.* comprise more than 1,000 individual measurements, which increases the overall precision by more than 30-fold, compared with a conventional 'snapshot'. Moreover, each measurement was made for numerous stars in that cluster. Spatial scanning has previously been used by researchers from the same group to study single stars, several thousand light years away, that are exceptionally bright<sup>6,7</sup>, but Brown *et al.* are the first to apply this technique to faint stars in a cluster at these sort of distances. (NGC 6397 is about 7,500 light years, or 2,390 parsecs, from Earth.)

The authors then used the same spatial scanning technique to measure the position of non-cluster stars in the background star field with incredible precision, enabling them to determine the displacement of cluster stars relative to each non-cluster star. But these relative parallaxes must be put into an absolute frame of reference, and setting such a frame is a complicated task. To do this, Brown and colleagues required coarse estimates (accurate to  $\pm 15\%$  of the true value) of the parallaxes for the non-cluster stars. The authors obtained these by determining the type and size of each star, and then assigning each the mean physical properties of its class, from which its distance (and therefore its parallax) can be determined<sup>8</sup>.

The Gaia mission also uses a scanning technique to obtain the positions of target objects, but sets the absolute frame using a sample of quasars (point-like galaxies that are unfathomably far away) from across the entire night sky<sup>9,10</sup>. Brown and colleagues' frame of reference has systematic uncertainties that are distinct from those of Gaia, and it could therefore provide a direct, independent means of testing the Gaia reference frame if it were to be expanded to include more star clusters.

The highly precise, long-distance parallax measurements provided by Gaia are a leap forward for astronomy. But, as in all fields of science, precision is not the only source of uncertainty. It is also crucial to understand the systematic uncertainty that is associated with a reference frame, partly so that this parameter can be included in data analyses, but also to devise a better means of establishing the frame. Systematic uncertainties can be reduced only by the addition of fresh, independent information, such as that provided by Brown and co-workers. The work involved in establishing these safeguards can be tedious and is often overlooked, but it is the bedrock of scientific progress. ■

**Rachael Beaton** is in the Department of Astrophysical Sciences, Princeton University, Princeton, New Jersey 08544, USA.

e-mail: rbeaton@princeton.edu

1. Brown, T. M. *et al. Astrophys. J. Lett.* **856**, L6 (2018).
2. Gaia Collaboration. Preprint at <https://arXiv.org/abs/1804.09365> (2018).
3. Bessel, F. W. *Mon. Not. R. Astron. Soc.* **4**, 152–161 (1938).

4. Lindegren, L. *et al. Astron. Astrophys.* **595**, A4 (2016).
5. Reid, M. J. & Honma, M. *Annu. Rev. Astron. Astrophys.* **52**, 339–372 (2014).
6. Riess, A. G., Casertano, S., Anderson, J., Mackenty, J. & Filippenko, A. V. *Astrophys. J.* **785**, 161 (2014).

7. Casertano, S. *et al. Astrophys. J.* **825**, 11 (2016).
8. Benedict, G. F. *et al. Publ. Astron. Soc. Pacif.* **106**, 327–336 (1994).
9. Gaia Collaboration. Preprint at <https://arXiv.org/abs/1804.09377> (2018).
10. Gaia Collaboration. *Astron. Astrophys.* **595**, A1 (2016).

## DEVELOPMENTAL BIOLOGY

# Human embryonic stem cells get organized

**An embryo's body plan is established by a structure called the organizer. Evidence of this structure in humans has been lacking, but a stem-cell-based protocol has now enabled researchers to demonstrate its existence. SEE LETTER P.132**

OLIVIER POURQUÉ

In 1924, Hilde Mangold and Hans Spemann performed what became one of the most famous experiments in developmental biology. They grafted various parts of a pigmented salamander embryo onto an unpigmented host embryo, and showed that one grafted region induced unpigmented cells from the host to form an extra embryo, resulting in a 'double embryo' reminiscent of conjoined twins<sup>1</sup> (Fig. 1a). The duo named the grafted region the organizer, because of its extraordinary ability to organize the host cells around it. But in the almost 100 years since this experiment, technical and ethical difficulties have prevented researchers from demonstrating the presence of an organizer in human embryos. On page 132, Martyn *et al.*<sup>2</sup> use stem cells to circumvent these challenges and provide the first experimental description of the human organizer.

To fully understand the importance of the organizer, we must go back to the earliest stages of embryonic development. In vertebrates, the fertilized egg rapidly divides to form a ball of poorly organized cells. At a particular developmental time point, some cells on the surface of this ball become internalized, forming tissues called the endoderm and the mesoderm, which respectively give rise to the gut and to muscles and the skeleton. Other cells remain external and give rise to the skin and the nervous system. This fundamental process of internalization is called gastrulation.

The organizer lies immediately adjacent to the site at which cells become internalized during gastrulation. It gives rise to specific tissues lying along the midline of the embryo, including the notochord — a structure that controls aspects of development of the central nervous system and eventually contributes to the intervertebral discs. An equivalent of the salamander organizer has been found in fish

and birds and in mammals such as rodents<sup>3</sup>. In mammals, the structure that acts as an organizer is called the node because it resembles a knot, and the site of internalization is called the primitive streak.

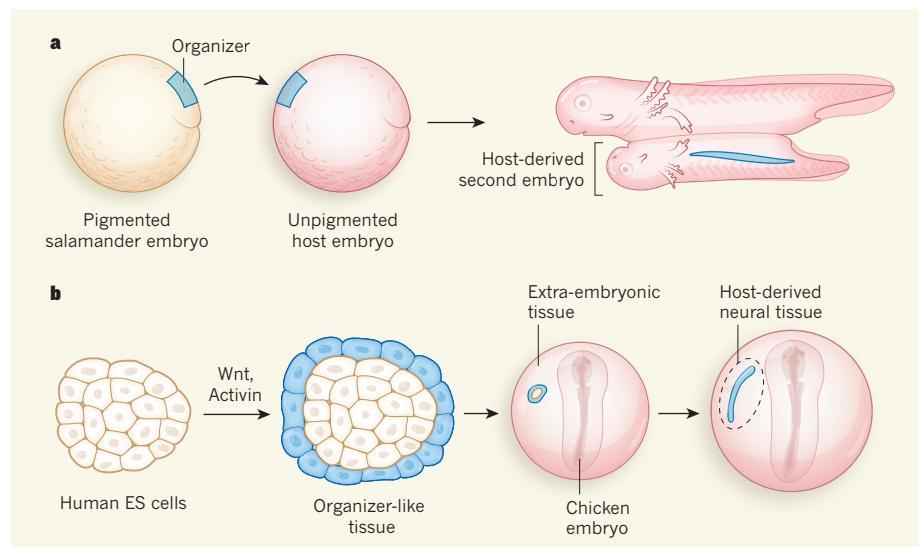
Unlike salamander embryos, mammals develop in the mother's womb. Accessing and culturing mammalian embryos is therefore difficult. Indeed, it wasn't until 1994 that grafts of a mouse node into a host embryo provided experimental proof of the existence of a structure that has organizer properties in mammals<sup>4</sup>. Although no perfect second embryos were formed in these experiments, the grafted nodes did induce the formation of host-derived neural tissues

and sometimes other embryonic tissues.

Human embryos greatly resemble mouse embryos and contain a structure that looks similar to the mouse node<sup>4</sup>. Theoretically, showing that this structure does indeed have the role of an organizer would require researchers to access embryos at three weeks of age (when gastrulation occurs), to graft the node onto a host embryo, and to test whether it induces the formation of a host-derived nervous system and skeletal structures. However, obtaining intact human embryos at this stage, for example from a pregnancy termination, is extremely problematic. Thus, whether the node represents a functional organizer in human embryos has remained unproven.

One alternative would be to let embryos obtained from *in vitro* fertilization (IVF) develop in culture until the three-week stage, when the node should be present. However, following an ethical consensus that is enshrined in law in many countries, human embryos cannot be cultured *in vitro* beyond 14 days, making these studies currently impossible.

A second alternative involves the use of pluripotent stem cells, which can give rise to all the body's cell lineages. Protocols to direct *in vitro* differentiation of these cells make it possible to recapitulate several aspects of



**Figure 1 | Experimental demonstration of organizer structures.** **a**, In 1924, an experiment<sup>1</sup> revealed the properties of an embryonic structure called the organizer. When taken from a pigmented salamander embryo and grafted onto an unpigmented host, the organizer induced the formation of a second embryo derived from unpigmented host cells. **b**, Martyn *et al.*<sup>2</sup> have demonstrated the existence of human cells endowed with similar properties, using human embryonic stem (ES) cells. The authors treated circular discs of ES cells with the growth-factor proteins Wnt and Activin to produce organizer-like cells (blue). When the discs are grafted onto the extra-embryonic tissue around a chick embryo, they induce the host tissue to form an elongated stretch of neural tissue — the standard test for organizer properties.

embryonic development in a culture dish.

Pluripotent cells derived from human embryos, called embryonic stem (ES) cells, generally form poorly organized colonies when grown in culture. However, the group that performed the current study previously induced<sup>5</sup> ES cells to self-organize in a way that resembles early embryonic development. They achieved this by culturing the cells on circular micropatterns (microprinted discs) of a material called extracellular matrix that is an optimal substrate for the cells) in the presence of the growth factor protein BMP4. The cultures formed endoderm and mesoderm but did not produce primitive streaks or node-like structures.

Martyn *et al.* took this strategy a step further. They successfully differentiated human ES cells into a node-like tissue by treating their micropatterned cultures with a combination of the growth factors Wnt and Activin, which are crucial for primitive streak and node formation in mice and other vertebrates<sup>6,7</sup>. This treatment led to the formation of a structure that showed characteristics of a primitive streak and to the induction of cells that produced organizer-specific proteins, such as Goosecoid<sup>8</sup>.

To test whether this structure also has the functional characteristics of an organizer, the authors grafted its cells onto chicken embryos, in an area destined to give rise to extra-embryonic tissues that support embryonic development. Remarkably, the grafted cells organized into a notochord-like tissue and induced host cells to form elongated neural tissue (Fig. 1b), demonstrating that the grafted structure has the properties of an organizer.

One could argue that these experiments still raise ethical concerns because they are performed using human ES cells derived from an early-stage human embryo. However, pluripotent cells generated by reprogramming adult cells, which have essentially identical properties to ES cells, could be used as an alternative, alleviating this concern in future studies.

Martyn and colleagues' experimental system provides an alternative to using embryos to study the human embryonic node. Moreover, their experiments suggest that there is striking evolutionary conservation of organizer function from fish to humans. How the organizer organizes the surrounding embryonic tissues into an embryo remains poorly understood, for now at least. But the ability to produce organizer tissue in unlimited amounts *in vitro* will allow researchers to dissect organizer function at an unprecedented level. ■

**Olivier Pourquié** is in the Department of Genetics, Harvard Medical School, and in the Department of Pathology, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA.  
e-mail: [pourquie@genetics.med.harvard.edu](mailto:pourquie@genetics.med.harvard.edu)

1. Spemann, H. & Mangold, H. *Wilhelm Roux Arch. Entw. Mech. Org.* **100**, 599–638 (1924).
2. Martyn, I., Kanno, T. Y., Ruza, A., Siggia, E. D. & Brivanlou, A. H. *Nature* **558**, 132–135 (2018).
3. Arias, A. M. & Steventon, B. *Development* **145**, dev159525 (2018).
4. Beddington, R. S. P. *Development* **120**, 613–620 (1994).
5. Warmflash, A., Sorre, B., Etoc, F., Siggia, E. D. &

6. Brivanlou, A. H. *Nature Meth.* **11**, 847–854 (2014).
7. Crease, D. J., Dyson, S. & Gurdon, J. B. *Proc. Natl Acad. Sci. USA* **95**, 4398–4403 (1998).
8. Gritsman, K., Talbot, W. S. & Schier, A. F. *Development* **127**, 921–932 (2000).
9. Cho, K. W. Y., Blumberg, B., Steinbeisser, H. & De Robertis, E. M. *Cell* **67**, 1111–1120 (1991).

This article was published online on 23 May 2018.

## CLIMATE CHANGE

# Tropical cyclones are becoming sluggish

The speed at which tropical cyclones travel has slowed globally in the past seven decades, especially over some coastlines. This effect can compound flooding by increasing regional total rainfall from storms. [SEE LETTER p.104](#)

CHRISTINA M. PATRICOLA

Tropical cyclones are among the deadliest and costliest of disasters ([go.nature.com/2h59avp](http://go.nature.com/2h59avp)), causing destruction not only from strong winds, but also from flooding and mudslides associated with storm surges and heavy rainfall. The total amount of storm rainfall over a given region can be extreme, regardless of the maximum storm wind speeds; it is proportional to the rainfall rate and inversely proportional to the translation speed<sup>1</sup> (how quickly a tropical cyclone passes over a region). Some studies have investigated trends in heavy rainfall from tropical cyclones over the past century<sup>2</sup> and future projections in tropical-cyclone rainfall rates<sup>3</sup>, but the translation speed has received less focus. On page 104, Kossin<sup>4</sup> investigates global trends in tropical-cyclone translation speed, and regional trends over individual ocean basins and adjacent land. He finds that translation speeds have slowed, suggesting that the total amount of regional rainfall from tropical cyclones might have increased.

Kossin analysed 68 years of observations made from 1949 to 2016, the longest period for which global data on the locations of tropical cyclones were available. The uncertainty associated with observed trends in translation speed is minimal during this period, because the locations of the tropical cyclones are accurately known. By contrast, it is more difficult to detect trends in the number and intensity of tropical cyclones during this period, because some of these cyclones were not detected in the pre-satellite era<sup>5</sup> (before the 1960s). Kossin finds a 10% global decrease in tropical-cyclone translation speed over this period, a trend that

withstands rigorous statistical testing and is dominated by tropical cyclones over the ocean.

The author found that changes in the translation speed of tropical cyclones over land — which are more relevant to society than those over the ocean — vary substantially by region. This is unsurprising, because only 10% of the original data are for such cyclones, and categorizing by region reduces the data sample further, making it more difficult to detect a signal among the noise. Nonetheless, statistically significant slowdowns of 20–30% have occurred over land regions next to the western North Pacific Ocean, the North Atlantic Ocean and around Australia.

Kossin's work highlights the importance of considering how global-scale atmospheric circulation can influence regional totals of tropical-cyclone rainfall. Tropical cyclones tend to 'go with the flow', meaning that the direction and speed at which they travel are guided by the winds in the surrounding environment. Therefore, any change in tropical circulation could conceivably affect tropical-cyclone translation speed, as Kossin reasons.

One limitation of this study is that it leaves open the question of what is happening to the rate of tropical-cyclone rainfall. The laws of thermodynamics reveal that, as the atmosphere warms by 1 °C, the amount of moisture it can hold increases by 7%. This suggests that global warming can enhance rainfall. However, it is unclear whether there are statistically robust trends in the total amount of regional tropical-cyclone rainfall, or how much the translation-speed slowdowns reported by Kossin could contribute to them. The availability and quality of data pose a challenge to our understanding of rainfall in general — the spatial distribution of rain gauges and radar observations of rainfall vary regionally, and satellite observations are limited to the past few decades and must be analysed using various assumptions to extract rainfall data. However,

**"The study finds a 10% global decrease in the speed at which tropical cyclones move."**



**Figure 1 | Hurricane Harvey seen from space.** The 2017 tropical cyclone known as Hurricane Harvey was particularly destructive, in part because it moved unusually slowly. Kossin<sup>4</sup> reports that the average speed with which tropical cyclones pass over a region has slowed since 1949.

if similar results are obtained from different data sources in overlapping periods, then any observed trends in rainfall can be considered to be robust.

Kossin's findings raise several questions, especially regarding 'stalled' tropical cyclones, which can be particularly destructive. Such cyclones are characterized by having an extremely slow translation speed (such as Typhoon Morakot<sup>1</sup>, which moved over Taiwan with a translation speed as slow as 5 kilometres per hour in 2009), a track that recurves or loops over a region more than once (such as Cyclone Hyacinthe, which looped past the island of Réunion three times in 1980), or both (such as Hurricane Harvey, which meandered along the coast of Texas in 2017; Fig. 1). Kossin reports that the probability of tropical cyclones having translation speeds slower than  $20 \text{ km h}^{-1}$  is significantly greater in the latter half of the observation period. However, it is not known whether stalled cyclones have become more frequent, nor how natural variability and anthropogenic climate change might contribute to such a trend. It is also unclear whether the incidence of stalled tropical cyclones will change in the future.

As Kossin points out, part of the challenge in understanding variability and change in the occurrence of stalled tropical cyclones lies in the lack of a quantitative metric. Moreover, stalled tropical cyclones are relatively rare, making it difficult to evaluate whether

there are statistically significant trends in the limited observations available. Statistical methods can help to quantify trends, but are sometimes less suitable for understanding the physical drivers.

Dynamic global climate models offer another solution to the problem of understanding stalled tropical cyclones. Computational simulations can represent current and future climates by changing the atmospheric concentrations of greenhouse gases and aerosols in such models. Dynamic models can also be used to separate the influences of natural variability and anthropogenic change. Advances in supercomputing now allow more global-climate simulations producing tropical-cyclone-like features than was previously possible. Collaborations between scientists studying tropical cyclones and those performing high-resolution climate simulations are thus producing valuable data sets<sup>6,7</sup>, even though the climate models are imperfect. Computer software has been developed that quickly identifies tropical cyclones and their characteristics within the petabytes of model data generated by these efforts<sup>8</sup>. And although low-resolution global climate models represent tropical cyclones poorly, statistical-dynamic models<sup>9,10</sup> have been developed that use ocean and atmospheric states produced by such models as inputs for simulating tropical cyclones at low computational cost.

To strengthen the resilience of coastal and

island communities to tropical cyclones, it is crucial to quantify and understand variability and change, not only in the number of tropical cyclones for different ocean basins, but also in the characteristics of tropical cyclones, including translation speed and its links with rainfall totals. Kossin's work paves the way towards developing this understanding, and raises questions that scientists can address using combinations of observations and modelling, to balance the benefits and limitations of each type of approach. ■

**Christina M. Patricola** is in the *Climate and Ecosystems Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA.*

*e-mail: cmpatricola@lbl.gov*

- Chien, F.-C. & Kuo, H.-C. *J. Geophys. Res. Atmos.* **116**, D05104 (2011).
- Kunkel, K. E. *et al. Geophys. Res. Lett.* **37**, L24706 (2010).
- Walsh, K. J. E. *et al. WIREs Clim. Change* **7**, 65–89 (2016).
- Kossin, J. P. *Nature* **558**, 104–107 (2018).
- Landsea, C. W., Harper, B. A., Hoarau, K. & Knaff, J. A. *Science* **313**, 452–454 (2006).
- Walsh, K. J. E. *et al. Bull. Am. Meteorol. Soc.* **96**, 997–1017 (2015).
- Haarsma, R. J. *et al. Geosci. Model Dev.* **9**, 4185–4208 (2016).
- Prabhat *et al. Proc. Comput. Sci.* **9**, 866–876 (2012).
- Emanuel, K. J. *Clim.* **19**, 4797–4802 (2006).
- Lee, C.-Y., Tippett, M. K., Sobel, A. H. & Camargo, S. J. *J. Adv. Model. Earth Syst.* **10**, <https://doi.org/10.1002/2017MS001186> (2018).

# Motion processing picks up speed in the brain

Recordings of individual neurons in the mouse brain reveal a main mechanism for motion processing in the primary visual cortex. These findings are likely to have implications for other species. [SEE ARTICLE P.80](#)

JOSE MANUEL ALONSO

Imagine life without visual motion. You can easily recognize a face, but you cannot safely cross the street because you do not see cars moving. Instead, you see snapshots of stationary cars that change position at unpredictable times. Most of us do not have this problem, because healthy brains contain many neurons that are direction selective — they fire strongly when they detect movement in one direction (known as their preferred direction, towards the left of the eye's field of view, for instance), and relatively weakly in response to movement in the opposite direction<sup>1</sup>. These direction-selective neurons are found in different parts of the brain, including the cerebral cortex. On page 80, Lien and Scanziani<sup>2</sup> report a major advance in our understanding of how neurons in the brain's primary visual cortex become direction selective.

Sensory information is transmitted from the eye to the primary visual cortex through a neuronal structure called the thalamus. Thalamic and cortical neurons respond only to stimuli within a small portion of the field of view, known as their receptive field. Previous studies<sup>3–7</sup> in cats suggest that direction selectivity in a cortical neuron arises from the combined activity of multiple thalamic neurons, which converge on the cortical neuron. These thalamic inputs have overlapping receptive fields, do not have direction selectivity themselves, and are temporally diverse in terms of their responses to stimuli. In this model, different thalamic inputs responding to moving stimuli over differing time periods generate a preferred direction of movement in the cortical neuron. However, finding direct experimental evidence for this model has been technically challenging.

Lien and Scanziani have overcome the technical limitations of the past thanks to the emergence of powerful tools that enable researchers to genetically manipulate neurons and measure the activity of multiple thalamic inputs to one cortical neuron. First, the authors used genetic-engineering techniques to silence the non-thalamic inputs to the primary visual cortex of mice, so that they could isolate the effects of thalamic inputs on direction selectivity in cortical neurons. These experiments replicated previous findings in cats<sup>4</sup>,

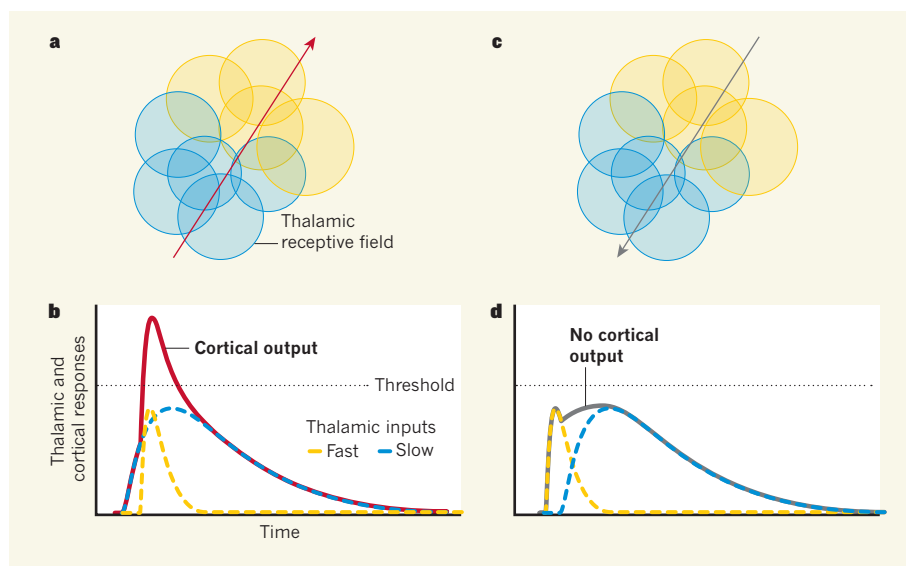
confirming that cortical cells maintain their direction selectivity when non-thalamic inputs are inactivated. Thus, direction selectivity in the primary visual cortex of mice originates at thalamic inputs.

Next, the researchers investigated the neuronal mechanisms underlying cortical direction selectivity in mice. In cats and primates<sup>8–10</sup>, some regions of the cortical receptive field respond more slowly and less transiently to visual stimuli than do other regions. The cortical neurons are preferentially activated by stimuli moving from slow- to fast-responding regions, because these stimuli produce the maximum neuronal response through temporal synchronization. To understand this synchronization, imagine two people who want to clap at the same time but who, when called to clap, respond at different rates — one more slowly than the other. To synchronize the

claps for maximum impact, the slower clapper has to receive the signal first, and the faster clapper, second. As in cats and primates, Lien and Scanziani found that cortical neurons in mice prefer a moving direction that goes from slow to fast regions of the receptive field.

When the authors silenced non-thalamic inputs, each cortical neuron still had slow and fast regions within its receptive field and preferred moving stimuli that crossed the slow region first. This finding shows, for the first time, that thalamic inputs are sufficient to generate both slow and fast responses in different regions in the cortical receptive field, and that these regions generate direction selectivity through synchronization. The authors also demonstrated that the temporal differences between the responses of different cortical receptive-field regions were most pronounced at the end of a response to a stimulus (the response decay): activity generated by slow regions decayed more gradually than activity generated by fast regions. These findings are consistent with work in cats and primates<sup>7,9</sup> that suggests that cortical direction selectivity is reliant mainly on synchronization of the response decay.

In their final and most impressive experiments, Lien and Scanziani recorded simultaneously the activity of a single cortical neuron and its thalamic inputs. The beautiful data collected in these experiments provide the strongest evidence so far that the slow and fast regions of the cortical receptive field



**Figure 1 | Generating direction selectivity in the brain's primary visual cortex.** Neurons in the primary visual cortex respond to movements in a particular direction within a small portion of the eye's field of view — their receptive field. Lien and Scanziani<sup>2</sup> have revealed a main mechanism for direction selectivity in this brain region in mice. **a**, Neurons in the brain's thalamus that have overlapping receptive fields send inputs to a single cortical neuron. Their receptive fields can respond to movement in either a fast, transient manner (yellow) or a slow, sustained manner (blue). The authors demonstrate that cortical neurons are preferentially activated by movement in the direction (red arrow) that causes slow thalamic inputs to respond first, followed by fast inputs. **b**, As the graph shows, this is because signals from both types of input are temporally synchronized, maximizing the cumulative cortical response (red line), which crosses the threshold required to generate an electrical output. **c**, By contrast, movement in the opposite direction (grey arrow) activates the fast thalamic inputs first. **d**, This generates a weaker cortical response (grey line), which does not cross the threshold because the thalamic inputs are less synchronized, and so does not produce an electrical output.

are generated by multiple thalamic inputs that have temporally different responses to the stimulus (Fig. 1). Thalamic inputs that respond slowly to visual stimuli generate slow responses in cortical regions, whereas those responding faster generate fast responses.

Lien and Scanziani's results, taken together with previous work<sup>3–10</sup>, raise the interesting possibility that cortical direction selectivity is generated through a common mechanism — the convergence of temporally diverse thalamic inputs — in rodents, cats and primates. But as with all research, some questions remain open.

For instance, the authors focus their study on the middle layers of the visual cortex, which receive the bulk of the thalamic input<sup>11</sup>. As Lien and Scanziani show, many thalamic inputs in these middle cortical layers are not direction selective, but their combined activity is. It remains unclear whether thalamic inputs that target other cortical layers (or serve other functions) can encode direction selectivity through different mechanisms. For example, neurons in the superficial layers of the cortex might derive their direction selectivity from thalamic neurons that are themselves direction selective<sup>12</sup>.

It is also known that thalamic inputs to the visual cortex are arranged by their receptive-field position — inputs that have receptive fields close to one another in the field of view are clustered together. However, it is not yet known whether the thalamic inputs are also arranged according to their temporal properties. If so, this could explain why spatial position and direction preference tend to change together in different

neurons across the visual–cortical map<sup>13</sup>.

Whatever the answers are, it is becoming increasingly clear that the visual cortex generates stimulus selectivity, such as preferences for direction and orientation, through thalamo–cortical convergence. Lien and Scanziani's work shows that this mechanism is better preserved across mammals than was previously thought. ■

**Jose Manuel Alonso** is in the College of Optometry, State University of New York, New York, New York 10036, USA.  
e-mail: [jalonso@sunyopt.edu](mailto:jalonso@sunyopt.edu)

- Hubel, D. H. & Wiesel, T. N. *J. Physiol.* **160**, 106–154 (1962).
- Lien, A. D. & Scanziani, M. *Nature* **558**, 80–86 (2018).
- Alonso, J.-M., Usrey, W. M. & Reid, R. C. *J. Neurosci.* **21**, 4002–4015 (2001).
- Ferster, D., Chung, S. & Wheat, H. *Nature* **380**, 249–252 (1996).
- Saul, A. B. & Humphrey, A. L. *J. Neurophysiol.* **64**, 206–224 (1990).
- Saul, A. B. & Humphrey, A. L. *J. Neurophysiol.* **68**, 1190–1208 (1992).
- Stanley, G. B. et al. *J. Neurosci.* **32**, 9073–9088 (2012).
- Reid, R. C., Soodak, R. E. & Shapley, R. M. *J. Neurophysiol.* **66**, 505–529 (1991).
- Livingstone, M. S. *Neuron* **20**, 509–526 (1998).
- McLean, J. & Palmer, L. A. *Vision Res.* **29**, 675–679 (1989).
- Lorente de No, R. In *Physiology of the Nervous System* (ed. Fulton, J.) 291–340 (Oxford Univ. Press, 1938).
- Cruz-Martín, A. et al. *Nature* **507**, 358–361 (2014).
- Kremkow, J., Jin, J., Wang, Y. & Alonso, J. M. *Nature* **533**, 52–57 (2016).

This article was published online on 23 May 2018.

## ENGINEERING

# Two artificial synapses are better than one

**Emerging nanoelectronic devices could revolutionize artificial neural networks, but their hardware implementations lag behind those of their software counterparts. An approach has been developed that tips the scales in their favour. [SEE ARTICLE P.60](#)**

GINA C. ADAM

Inspired by the brain's neural networks, scientists have for decades tried to construct electronic circuits that can process large amounts of data. However, it has been difficult to achieve energy-efficient implementations of artificial neurons and synapses (connections between neurons). On page 60, Ambrogio et al.<sup>1</sup> report an artificial neural network containing more than 200,000 synapses that can classify complex collections of images. The authors' work demonstrates that hardware-based neural networks that use emerging nanoelectronic devices

can perform as well as can software-based networks running on ordinary computers, while consuming much less power.

Artificial neural networks are not programmed in the same way as conventional computers. Just as humans learn from experience, these networks acquire their functions from data obtained during a training process. Image classification, which involves learning and memory, requires thousands of artificial synapses. The states (electrical properties) of these synapses need to be programmed quickly and then retained for future network operation.

Nanoscale synaptic devices that have programmable electrical resistance, such



## 50 Years Ago

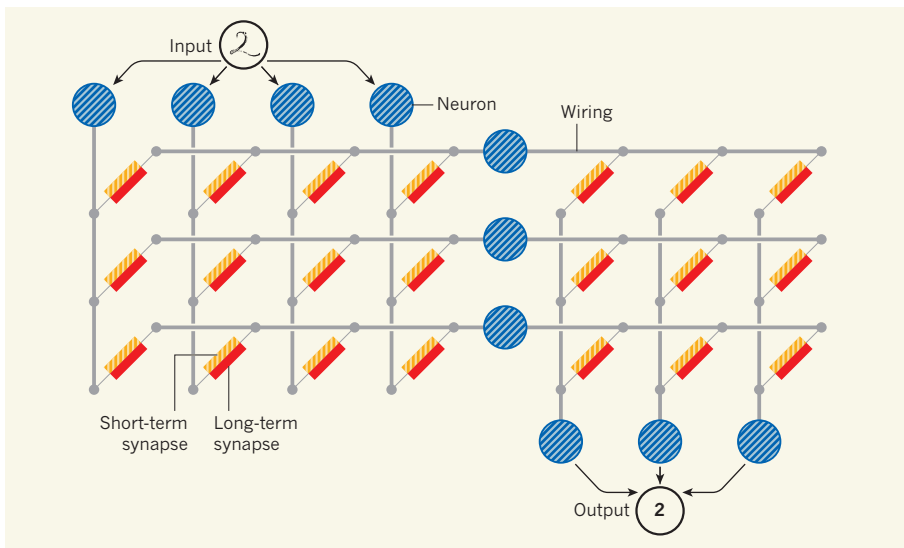
Reading aids for the blind have so far involved the use of intact sensory pathways and have progressed little beyond Braille and tape-recorded “talking-books”. Both these systems are quite expensive ... and both are slow in terms of information transfer to the reader ... At a recent meeting of the Physiological Society, Brindley and Lewin demonstrated a device for stimulating the visual cortex of man directly ... Essentially it consists of an array of radio receivers, encapsulated in silicone rubber and screwed to the skull ... Activation of a receiver stimulated the cortex: transmission was in the form of a train of short (200  $\mu$ s) pulses ... it does at least seem feasible to transmit visual information directly to the central visual pathways of the recently blind.

From *Nature* 8 June 1968

## 100 Years Ago

It happened last week that about 1 lb. of fresh lamb was put into an oven at night in order that it might be cooked by morning on the “hay-box” principle. It was in a casserole, with a little water. Similar treatment in the same oven on previous occasions had been very successful. At about 5 a.m. the casserole was examined, and the broth was found to be very well tasted, and the whole smelt fresh and good, but the meat when tested with a fork was not tender, and the fat (of which there was a good deal) was *entirely* unmelted. The casserole was returned to the oven (then quite cool) and taken out again after breakfast. The contents were then found to be smelling most offensively, as if extremely “high”. The fat was melted. The meat and broth were judged quite unfit for human food. I wonder if any of your readers would explain this curious development.

From *Nature* 6 June 1918



**Figure 1 | An artificial neural network containing two types of synapse.** Ambrogio *et al.*<sup>1</sup> report a hardware-based artificial neural network that is trained to classify complex images, such as handwritten numbers, with an accuracy similar to that of a software-based network. The network consists of artificial neurons linked by wires to two types of artificial synapse (connections between neurons). Short-term synapses (which can retain alterations in their synaptic state for milliseconds) are used regularly during network training, whereas long-term synapses (with state retention of years) are used mainly for memory. The long-term synapses are physical devices, whereas the neurons and short-term synapses are simulated computationally (indicated by hatching).

as phase-change-memory (PCM) devices, show promise because of their small physical size and excellent retention properties. PCM devices contain a material known as a chalcogenide glass, which can switch reversibly between an amorphous phase (of high resistance) and a crystalline phase (of low resistance). The device's resistance state is programmed by crystallizing part of the material using local heating produced by an applied voltage. This state is retained long after the voltage has been removed, and further programming can be achieved by crystallizing other parts of the material.

Unfortunately, PCM devices can be programmed in only one direction: from high to low resistance, by changing from low to high crystallinity. To achieve the desired resistance state with good precision, sequences of hundreds of voltage pulses are required. If the desired state is overshoot, the chalcogenide glass must be completely reset to the amorphous phase and the step-by-step programming restarted. This shortcoming, combined with variations between devices caused by the manufacturing process, can slow or even prevent network training, as previous work by the group that performed the current study has shown<sup>2</sup>. As a result, the prototype networks that have been constructed using these devices<sup>3,4</sup> are impractical and have much lower image-classification accuracies than do software-based networks.

The breakthrough of Ambrogio and colleagues' work lies in a two-tier, bio-inspired approach. In biological neural networks, short-term changes in the states of synapses support a variety of computations, whereas long-term

changes provide a platform for learning and memory<sup>5</sup>. For this reason, the authors' artificial neural network uses synaptic 'cells' that contain two types of synapse: short-term and long-term (Fig. 1).

The short-term synapses are used regularly during network training. They require only brief state retention, but fast and precise programming to the desired state. Such features are provided by an electronic switch called a transistor, which has a capacitor (a device for storing electric charge) attached to one of its electrodes, known as the gate<sup>6</sup>. The transistor's state is programmed by a fast voltage pulse applied to the gate. The capacitor maintains this voltage for a few milliseconds, providing brief state retention.

After the network has been trained on several thousand images and the short-term synapses have changed states substantially, the synaptic states are written into long-term synapses. The cycle is then repeated until all of the training images have been presented to the network. The long-term synapses are used for network operation after training is complete. They consist of PCM devices that have state-retention times of years, at the expense of tedious, energy-intensive programming.

An advantage of this technique is that the transfer of states from short- to long-term synapses can be done in electronic-circuit blocks separate from the network, while the network carries out other tasks. Moreover, although the authors' synaptic cells are more complicated in practice — containing one capacitor, two PCM devices and five transistors — they are still about half the size of artificial synapses used in other networks<sup>6</sup>.

Ambrogio *et al.* tested their synaptic-cell approach using a fairly complex artificial neural network containing multiple layers of neurons and more than 200,000 PCM devices. The authors carried out classification tasks using three standard sets of images: greyscale handwritten numbers from the MNIST database<sup>7</sup>, and colour images from the CIFAR-10 and CIFAR-100 databases<sup>8</sup>. The accuracies obtained were 98%, 88% and 68%, respectively. These results are strikingly similar to those obtained using TensorFlow, a leading neural-network software (see [www.tensorflow.org](http://www.tensorflow.org)).

Despite these impressive findings, a key limitation of the work is that only the PCM devices were actually fabricated; the other components of the synaptic cells and the neurons were simulated computationally. The authors took care to use accurate models that consider variations between transistors, and they proposed a method to minimize the impact of such variability on synaptic-cell performance. Most importantly, they carried out a detailed power assessment, and found that their proposed technology would consume about 100 times less power than current state-of-the-art networks, while providing a similar classification performance. Nevertheless, only a working hardware prototype will convince industry of the technology's performance and low-power advantages. Furthermore, the estimated power consumption is still a far cry from that of biological neural networks, leaving plenty of room for improvement.

However, Ambrogio and colleagues' work is more than a crucial stepping stone to the integration of PCM devices in neural-network hardware. It will also inspire device research, because it creates a need for nanoscale short-term synapses to replace the bulky transistor-capacitor ones. A wall in emerging memory technologies has been breached — networks based on these devices can work as well as do their software counterparts. This finding suggests that advances in artificial intelligence will not only continue, but also be accelerated by emerging hardware. ■

**Gina C. Adam** is at the National Institute for Research and Development in Microtechnologies, Bucharest 077190, Romania.  
e-mail: [gina.adam@imt.ro](mailto:gina.adam@imt.ro)

1. Ambrogio, S. *et al.* *Nature* **558**, 60–67 (2018).
2. Burr, G. W. *et al.* 2014 IEEE Int. Electron Devices Meet. 29.5.1–29.5.4 (IEEE, 2014).
3. Gokmen, T. & Vlasov, Y. *Front. Neurosci.* **10**, 333 (2016).
4. Yu, S. *et al.* 2015 IEEE Int. Electron Devices Meet. 17.3.1–17.3.4 (IEEE, 2015).
5. Abbott, L. F. & Regehr, W. G. *Nature* **431**, 796–803 (2004).
6. Kim, S., Gokmen, T., Lee, H.-M. & Haensch, W. E. 2017 IEEE 60th Int. Midwest Symp. Circuits Systems 422–425 (IEEE, 2017).
7. Lecun, Y., Bottou, L., Bengio, Y. & Haffner, P. *Proc. IEEE* **86**, 2278–2324 (1998).
8. Krizhevsky, A. *Learning Multiple Layers of Features From Tiny Images* Ch. 3; [www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf](http://www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf); <https://www.cs.toronto.edu/~kriz/cifar.html> (2009).

# The many possible climates from the Paris Agreement's aim of 1.5 °C warming

Sonia I. Seneviratne<sup>1\*</sup>, Joeri Rogelj<sup>1,2,3,4</sup>, Roland Séférian<sup>5</sup>, Richard Wartenburger<sup>1</sup>, Myles R. Allen<sup>3</sup>, Michelle Cain<sup>3</sup>, Richard J. Millar<sup>3</sup>, Kristie L. Ebi<sup>6</sup>, Neville Ellis<sup>7</sup>, Ove Hoegh-Guldberg<sup>8</sup>, Antony J. Payne<sup>9</sup>, Carl-Friedrich Schleussner<sup>10,11,12</sup>, Petra Tschakert<sup>7</sup> & Rachel F. Warren<sup>13</sup>

**The United Nations' Paris Agreement includes the aim of pursuing efforts to limit global warming to only 1.5 °C above pre-industrial levels. However, it is not clear what the resulting climate would look like across the globe and over time. Here we show that trajectories towards a '1.5 °C warmer world' may result in vastly different outcomes at regional scales, owing to variations in the pace and location of climate change and their interactions with society's mitigation, adaptation and vulnerabilities to climate change. Pursuing policies that are considered to be consistent with the 1.5 °C aim will not completely remove the risk of global temperatures being much higher or of some regional extremes reaching dangerous levels for ecosystems and societies over the coming decades.**

Since 2010, international climate policy under the United Nations moved the public discourse from a focus on atmospheric concentrations of greenhouse gases to a focus on distinct global temperature targets above the pre-industrial period<sup>1,2</sup>. In 2015, this led to the inclusion of a long-term temperature goal in the Paris Agreement<sup>1</sup> that makes reference to two levels of global mean temperature increase: 1.5 °C and 2 °C. The former is set as an ideal aim ("pursuing efforts to limit the temperature increase to 1.5 °C") and the latter is set as an upper bound ("well below 2 °C")<sup>1</sup>. This change in emphasis allows a better link between mitigation targets and the required level of adaptation ambition<sup>3,4</sup>.

Assessing the effects of the reduction of anthropogenic forcing through a single qualifier—namely, global mean temperature change compared with the pre-industrial climate—however, also entails risks. This deceptively simple characterization may lead to an oversimplified perception of human-induced climate change and of the potential pathways to limit the impacts of greenhouse gas forcing. We highlight here the multiple ways in which a 1.5 °C global warming may be realized. These alternative '1.5 °C warmer worlds' are related to (a) the temporal and regional dimension of 1.5 °C pathways, (b) model-based spread in regional climate responses, (c) climate noise and (d) a range of possible options for mitigation and adaptation. We also highlight potential high-risk temperature outcomes of mitigation pathways currently considered to be consistent with 1.5 °C owing to uncertainties in relating greenhouse gas emissions to subsequent global warming, and to uncertainties in relating global warming to associated regional climate changes.

## Definition of a '1.5 °C warming'

Global mean temperature is a construct: it is the globally averaged temperature of Earth that can be derived from point-scale ground observations or computed in climate models. Global mean temperature is defined over a given time frame (for example, averaged over a month, a year or multiple decades). As a result of climate variability, which is due to internal variations of the climate system and temporary naturally induced forcings (for example, from volcanic eruptions), a

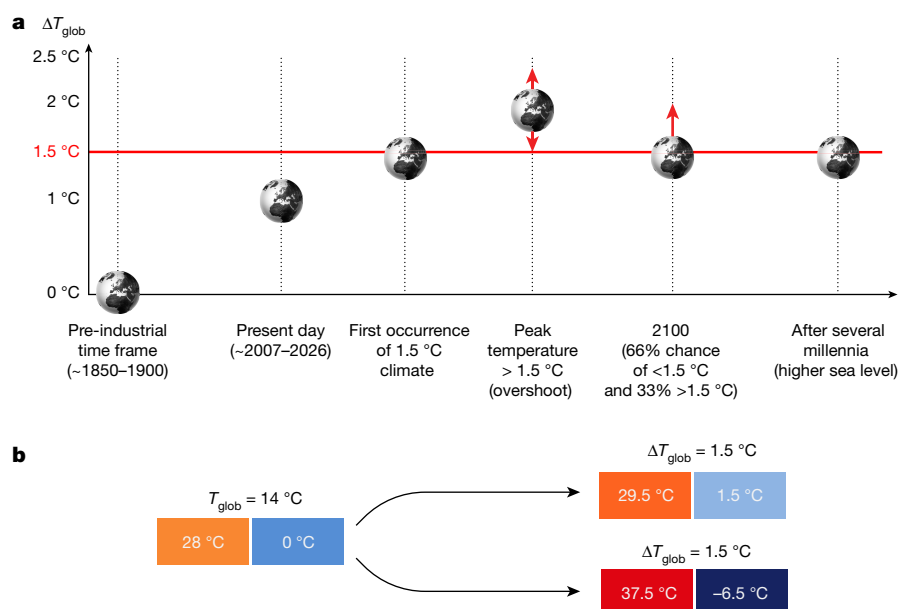
climate-based global mean temperature typically needs to be defined over several decades (at least 30 years according to the World Meteorological Organization)<sup>5</sup>. Hence, to determine a 1.5 °C global temperature warming, one needs to agree on a reference period (assumed here to be 1850–1900 inclusive, unless otherwise indicated), and on a time frame over which a 1.5 °C mean global warming is observed (assumed here to be of the order of one to several decades). Comparisons of global mean temperatures from models and observations are also not straightforward: not all points over Earth's surface are continuously observed, leading to methodological choices about how to deal with data gaps<sup>6</sup> and with the mixture of air temperature over land and surface water temperatures of oceans<sup>7</sup> when comparing full-field climate models with observational products.

## Temporal and spatial dimensions

There are two important temporal dimensions of 1.5 °C warmer worlds: (a) the time period over which the 1.5 °C warmer climate is assessed; and (b) the pathway followed before reaching this temperature level, in particular whether global mean temperature returns to the 1.5 °C level after previously exceeding it for some time (also referred to as 'overshooting'; see Fig. 1a). As highlighted hereafter, for some components of the coupled human–Earth system, there are substantial differences in risk between 1.5 °C of warming in the year 2040, 1.5 °C of warming in 2100 either with or without earlier overshooting, and 1.5 °C warming after several millennia at this warming level.

The time period over which 1.5 °C warming is reached is relevant because some slow-varying elements of the climate system respond with a delay to radiative forcing and to the associated temperature anomalies. Hence the status of such slow-varying elements will change over time, even if the warming is stabilized at 1.5 °C over several decades, centuries or millennia. This is the case with the melting of glaciers, ice caps and ice sheets and their contribution to future sea level rise, as well as the warming and expansion of the oceans, so that a substantial component of contemporary sea-level rise is a response to past warming. In

<sup>1</sup>Institute for Atmospheric and Climate Science, ETH Zurich, Zurich, Switzerland. <sup>2</sup>International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria. <sup>3</sup>Environmental Change Institute, School of Geography and the Environment, University of Oxford, Oxford, UK. <sup>4</sup>Grantham Institute, Imperial College London, London, UK. <sup>5</sup>Centre National de Recherches Météorologiques, Météo-France/CNRS, Toulouse, France. <sup>6</sup>Department of Global Health, University of Washington, Seattle, WA, USA. <sup>7</sup>School of Agriculture and Environment, University of Western Australia, Perth, Western Australia, Australia. <sup>8</sup>Global Change Institute, University of Queensland, Brisbane, Queensland, Australia. <sup>9</sup>University of Bristol, Bristol, UK. <sup>10</sup>Climate Analytics, Berlin, Germany. <sup>11</sup>IRITHESys, Humboldt University, Berlin, Germany. <sup>12</sup>Potsdam Institute for Climate Impact Research, Potsdam, Germany. <sup>13</sup>Tyndall Centre for Climate Change, School of Environmental Sciences, University of East Anglia, Norwich, UK. \*e-mail: [sonia.seneviratne@ethz.ch](mailto:sonia.seneviratne@ethz.ch)



**Fig. 1 | Temporal and spatial dimensions of 1.5 °C warmer worlds.**

**a**, Typical pathways of Earth's climate towards stabilization at 1.5 °C warming. Pre-industrial climate conditions are the reference for the determined global warming. Present-day warming corresponds to 1 °C compared to pre-industrial conditions. All emissions pathways compatible with 1.5 °C warming that are available in the literature<sup>12–15</sup> include overshooting over 1.5 °C warming before stabilization or further decline. We here illustrate the example of temperature stabilization at 1.5 °C in the long term, but temperatures could also decline further below 1.5 °C. **b**, Not

all conceivable 1.5 °C warmer climates are equivalent. These conceptual schematics illustrate the importance of the spatial dimension of distributed impacts associated with a given global warming, in the example of a simplified world with two surfaces of equal area (the given temperature anomalies are chosen for illustrative purposes and do not refer to specific 1.5 °C scenarios). Left, a reference world (without warming); top right, a world with 1.5 °C mean global warming that is equally distributed on the two surfaces; bottom right, a world with 1.5 °C mean global warming with high differences in regional responses.

addition, the rate of warming is also an important element of imposed stress for resulting risks, because it may affect adaptation or lack thereof<sup>8–10</sup>. For example, the faster the rate of change the fewer taxa (and hence ecosystems) can disperse naturally to track their climate envelope across Earth's surface<sup>8,11</sup>. Similarly, in human systems, faster rates of change in climate variables such as sea level rise present increasing challenges to adaptation to the point where attempts may be increasingly overwhelmed.

Whether mean global temperature temporarily overshoots the 1.5 °C limit is another important consideration. All currently available mitigation pathways projecting less than 1.5 °C global warming by 2100 include some probability of overshooting this temperature, with some time period during the twenty-first century in which warming higher than 1.5 °C is projected with a probability<sup>12–15</sup> of greater than 50%. This is inherent to the difficulty of limiting warming to 1.5 °C given that Earth at present is already very close to this warming level (about 1 °C warming for the current time frame relative to 1851–1900<sup>16</sup>). The implications of overshooting are essential for projecting future risks and for considering potentially long-lasting and irreversible impacts in the time frame of the current century and beyond, for instance associated with ice melting<sup>17</sup> and resulting sea level rise, loss of ecosystem functionality and increased risks of species extinction<sup>11</sup>, or loss of livelihoods, identity and sense of place and belonging<sup>18</sup>. Overshooting might cause some impact thresholds to be temporarily exceeded. This might be sufficient to cause permanent loss of ecosystems, or those systems and species able to adapt rapidly enough to cope with a particular rate of change would be faced with the challenge of adapting again to a lower level of warming post-overshoot. The chronology of emission pathways and their implied warming is also important for the more slowly evolving parts of the Earth system, such as those associated with sea level rise (see above). The remaining carbon budget available for emissions is very small, implying that deeper global mitigation efforts are required immediately if the duration and magnitude of the overshoot (exceeding the 1.5 °C level of warming) is to be minimized; see below and Table 1 and Box 1.

The spatial dimension of 1.5 °C warmer worlds is also important. Two worlds with similar global mean temperature anomalies may have very different risks depending on how the associated regional temperature anomalies are distributed (Fig. 1b). Differential geographical responses in temperature are induced by: (a) spatially varying radiative forcing (for example, associated with land use<sup>19–21</sup> or aerosols<sup>22</sup>; (b) differential regional feedbacks to the applied radiative forcing (for example, associated with soil moisture, snow, or ice feedbacks<sup>4,23</sup>); and (c) regional climate noise<sup>24</sup> (for example, associated with modes of variability or atmospheric weather variability). Similar considerations apply to regional changes in precipitation means and extremes, which are not globally homogeneous<sup>3,4</sup>. These regional temperature and precipitation anomalies and their rates of change determine the regional risks to human and natural systems and the challenges to adaptation which they face.

We note that mitigation, adaptation and development pathways may result in spatially varying radiative forcing. Although the greenhouse gases are well mixed, changes in land use or air pollution may strongly affect regional climate. Land-use changes can be associated, for example, with the implementation of increased bioenergy plantations<sup>25</sup>, afforestation, reforestation, or deforestation, and their resulting impacts on local albedo or evapotranspiration. Levels of aerosol concentrations may vary as a result of decreased air pollution<sup>22</sup>. Considering these regional forcings is essential when evaluating regional impacts, although there is still little available literature for 1.5 °C warmer worlds, or low-emissions scenarios in general<sup>22,26–28</sup>. The spatial dimension of regional climates associated with a global warming of 1.5 °C is also crucial when assessing risks associated with proposed climate engineering schemes based on solar radiation management (SRM, see below). Besides the geographical distribution of changes in climate, non-temperature-related changes are important, particularly where atmospheric CO<sub>2</sub> has additional and serious impacts through phenomena such as ocean acidification.

### Uncertainties of emissions pathways

Emissions pathways that are currently considered to be compatible with limiting global warming to 1.5 °C<sup>12–15</sup> are selected on the basis of their

**Table 1 | Description of different 1.5 °C and 2 °C warmer worlds**

		SCEN_1p5C emissions pathways currently considered compatible with a 66% chance of keeping warming below 1.5 °C in 2100 (allowing for a higher peak in temperature earlier)		SCEN_2C emissions pathways currently considered compatible with a 66% chance of keeping warming below 2 °C during the entire twenty-first century	
		'Probable' (66th percentile) outcome <sup>a</sup>	'Worst-case' 10% (90th percentile) outcome <sup>b</sup>	'Probable' (66th percentile) outcome <sup>a</sup>	'Worst-case' 10% (90th percentile) outcome <sup>b</sup>
<b>General characteristics of pathway</b>	Overshoot 1.5 °C in twenty-first century with >50% likelihood <sup>c</sup>	<b>Yes (13/13)</b>	<b>Yes (13/13)</b>	<b>Yes (10/10)</b>	<b>Yes (10/10)</b>
	<b>Overshoot 2 °C in twenty-first century with &gt;50% likelihood</b>	<b>No (0/13)</b>	<b>Yes (10/13)</b>	<b>No (0/10)</b>	<b>Yes (10/10)</b>
	Cumulative CO <sub>2</sub> emissions up to peak warming (relative to 2016) <sup>d</sup> (Gt CO <sub>2</sub> )	720 (650–750)	690 (650–710)	1,050 (1,020–1,140)	1,040 (930–1,140)
	Cumulative CO <sub>2</sub> emissions up to 2100 (relative to 2016) <sup>d</sup> (Gt CO <sub>2</sub> )		320 (200–340)		1,030 (910–1,140)
	Global greenhouse gas emissions in 2030 <sup>d</sup> (GtCO <sub>2</sub> yr <sup>-1</sup> )		22 (19–31)		28 (24–30)
<b>Possible climate range at peak warming (regional + global)</b>	Years of global net zero CO <sub>2</sub> emissions <sup>d</sup>		2070 (2067–2074)		2088 (2085–2092)
	<b>Global mean temperature anomaly at peak warming (°C)</b>	<b>1.75 (1.65–1.81)</b>	<b>2.13 (2.0–2.2)</b>	<b>1.93 (1.9–1.94)</b>	<b>2.44 (2.43–2.46)</b>
	Warming in the Arctic <sup>e</sup> , T <sub>night,min</sub> (°C)	5.04 (4.45–5.66)	6.29 (5.47–7.21)	5.70 (4.90–6.53)	7.25 (6.51–8.24)
	Warming in the contiguous USA <sup>e</sup> , T <sub>day,max</sub> (°C)	2.57 (2.04–2.95)	3.09 (2.71–3.58)	2.83 (2.34–3.27)	3.63 (3.23–3.98)
	Warming in Central Brazil <sup>e</sup> , T <sub>day,max</sub> (°C)	2.74 (2.39–3.22)	3.34 (3.05–3.92)	3.01 (2.62–3.50)	3.82 (3.44–4.15)
<b>Possible climate range in 2100 (regional + global)</b>	Drying in the Mediterranean region <sup>e</sup> (std <sup>f</sup> )	–1.27 (–2.43 to –0.45)	–1.40 (–2.64 to –0.52)	–1.14 (–2.18 to –0.50)	–1.42 (–2.74 to –0.67)
	Increase in heavy precipitation events <sup>f</sup> in Southern Asia <sup>e</sup> (%)	9.69 (6.79–14.90)	12.87 (7.90–22.78)	10.01 (6.97–17.11)	17.45 (10.15–24.03)
	<b>Global mean temperature warming in 2100 (°C)</b>	<b>1.44 (1.44–1.48)</b>	<b>1.88 (1.85–1.93)</b>	<b>1.89 (1.88–1.91)</b>	<b>2.43 (2.42–2.46)</b>
	Warming in the Arctic <sup>e</sup> , T <sub>night,min</sub> <sup>f</sup> (°C)	4.21 (3.65–4.71)	5.55 (4.80–6.35)	5.58 (4.82–6.38)	7.22 (6.49–8.16)
	Warming in the contiguous USA <sup>e</sup> , T <sub>day,max</sub> (°C)	2.03 (1.64–2.49)	2.73 (2.21–3.22)	2.76 (2.23–3.24)	3.64 (3.23–3.97)
	Warming in central Brazil <sup>e</sup> , T <sub>day,max</sub> <sup>f</sup> (°C)	2.25 (2.02–2.60)	2.92 (2.55–3.44)	2.94 (2.58–3.47)	3.80 (3.43–4.12)
	Drying in the Mediterranean region <sup>e</sup> (std <sup>f</sup> )	–0.96 (–1.94 to –0.28)	–1.09 (–2.16 to –0.48)	–1.10 (–2.15 to –0.46)	–1.41 (–2.69 to –0.64)
	Increase in heavy precipitation events <sup>f</sup> in Southern Asia <sup>e</sup> (%)	8.29 (4.52–11.98)	10.59 (6.75–16.64)	10.55 (6.83–16.64)	17.21 (10.24–24.03)

Data are based on scenarios currently considered compatible with 1.5 °C and 2 °C warming<sup>15</sup>, including projections of changes in regional climate associated with resulting global temperature levels derived following previous studies<sup>4,37</sup> (see Supplementary Information for corresponding estimates from scenarios assessed in the IPCC AR5<sup>12,14</sup> and for median estimates).

<sup>a</sup>66th percentile for global temperature (that is, 66% likelihood of being at or below values)

<sup>b</sup>90th percentile for global temperature (that is, 10% likelihood of being at or above values)

<sup>c</sup>All 1.5 °C scenarios include a substantial probability of overshooting above 1.5 °C global warming before returning to 1.5 °C.

<sup>d</sup>The values indicate the median with the interquartile range in parentheses (25th percentile and 75th percentile)

<sup>e</sup>The regional projections in these rows provide the median and the range [q25, q75] associated with the median global temperature outcomes of the considered mitigation scenarios at peak warming (see Box 1 and Supplementary Information).

<sup>f</sup>'std' indicates drying of soil moisture expressed in units of standard deviations of pre-industrial climate (1861–1880) variability (where –1 is dry; –2 is severely dry; and –3 is very severely dry);

<sup>R</sup><sub>max,5-day</sub> is the annual maximum consecutive 5-day precipitation.

<sup>g</sup>As for footnote e, but for the regional responses associated with the median global temperature outcomes of the considered mitigation scenarios in 2100 (see Box 1 and Supplementary Information for details).

probability of limiting warming to below 1.5 °C by 2100 given current knowledge of how the climate system is likely to respond. Typically, this probability is set at 50% or 66% (the chance of limiting warming in 2100 to 1.5 °C or lower). The adequacy of these levels of probability is more a political than a scientific question. This implies that even when diligently following such 1.5 °C pathways from today onwards, there is considerable probability that the 1.5 °C limit will be exceeded. This also includes some possibilities of warming being substantially higher than 1.5 °C (see discussion below of the 10% worst-case scenarios). These risks of alternative climate outcomes are not negligible and need to be factored into the decision-making process.

Table 1 provides an overview of the outcomes of emissions pathways that are currently considered 1.5 °C- and 2 °C-compatible with a specific probability<sup>15</sup> (and broadly consistent with the literature assessed in the Intergovernmental Panel on Climate Change Fifth Assessment Report (IPCC AR5)<sup>12,14</sup>, see Box 1 and Supplementary Information). Both 'probable' (66th percentile of scenarios, which remains below the respective temperature targets, that is, with two-thirds of the scenarios having lower or equal global warming) and 'worst-case' (90th percentile, that is, with 10% of scenarios having higher or equal global

warming) outcomes of these pathways are presented, including the resulting global temperatures and regional climate changes (see below and Box 1 for details, and the Supplementary Information for median outcomes). The reported net cumulative CO<sub>2</sub> emissions characteristics for these scenario categories include the effects of carbon dioxide removal options (also termed "negative emissions"<sup>29</sup>), which explains the decrease in cumulative CO<sub>2</sub> budgets after peak warming. Possible proposed carbon dioxide removal approaches include bioenergy use with carbon capture and storage (BECCS) or afforestation and changes in agricultural practice increasing carbon sequestration on land<sup>29</sup>. We note that the use of these approaches is controversial and could entail separate risks, for instance those related to competition for land use<sup>30,31</sup>. Their implementation is at present also still very limited, and the feasibility of their deployment as simulated in low-emissions scenarios has been questioned<sup>32</sup>. Current publications<sup>12,14,15</sup> indicate that scenarios in line with limiting year-2100 warming to below 1.5 °C require strong and immediate mitigation measures and would require some degree of carbon dioxide removal. Alternative scenario configurations can be considered to limit the amount of carbon dioxide removal<sup>32,33</sup>. The current scenarios<sup>15</sup> as well as recent publications<sup>34–36</sup> provide updated

## Box 1

## Emissions budgets and regional projections for 1.5 °C and 2 °C warming

The emissions budget estimates of Table 1 are based on scenarios currently considered compatible with limiting global warming ( $\Delta T_{\text{glob}}$ ) to 1.5 °C and 2 °C, either in 2100 or during the entire twenty-first century<sup>15</sup>. The emissions pathways are determined based on their probability of limiting  $\Delta T_{\text{glob}}$  below 1.5 °C or 2 °C by 2100 using the probabilistic outcomes of a simple climate model (MAGICC<sup>71</sup>) exploring the range of climate system response as assessed in the IPCC AR5<sup>72</sup>. The 50th (Supplementary Information), 66th and 90th percentile (Table 1) MAGICC global transient climate response values in the scenarios are 1.7 °C, 1.9 °C and 2.4 °C, respectively, which is consistent overall with the range assessed for this parameter (>66% in the 1–2.5 °C range, less than 5% greater than 3 °C) in the IPCC AR5<sup>72</sup>. The current airborne fraction (ratio of accumulated atmospheric CO<sub>2</sub> to CO<sub>2</sub> emissions over the decade 2011–2020) in these scenarios with this MAGICC version has been estimated at 0.55, which is 20% higher than the central estimate for the most recent decade given in refs<sup>73,74</sup>, but ref. 74 emphasizes that this quantity is uncertain and subject to variability over time. The estimates provided are consistent with corresponding values from scenarios assessed in the IPCC AR5<sup>12,14</sup> (see Supplementary Table 1), but have slightly larger estimates for the remaining cumulative CO<sub>2</sub> budgets, consistent with other recent publications<sup>34–36</sup>. Both sets of scenarios imply that for limiting  $\Delta T_{\text{glob}}$  below 1.5 °C by 2100 strong near-term mitigation measures are needed, supported by technologies capable of enabling net-zero global CO<sub>2</sub> emissions near to mid-century.

Table 1 and Figs. 2 and 3 also provide estimates of regional responses associated with given  $\Delta T_{\text{glob}}$  levels (at peak warming and in 2100). The values are computed based on decadal averages of 26 CMIP5 global climate model simulations and all four RCP scenarios following the approach from refs<sup>4,37</sup> (see Supplementary Information for more details). Decades corresponding to a 1.5 °C or 2 °C warming are those in which the last year of the decade reaches this temperature, consistent with previous publications<sup>3,4,37</sup>. Corresponding regional responses for the median estimates of the scenarios considered are provided in Supplementary Table 2 and Supplementary Figs. 1 and 2. The respective estimates of spread for recent (0.5 °C) and present-day (1 °C) global warming are provided in Supplementary Fig. 3.

Figure 4 is based on the same subset of the 26 CMIP5 models as was used for Table 1 and Figs. 2 and 3, but uses RCP8.5 simulations only. For each simulation, the ensemble percentiles are calculated for the time step corresponding to the decade at which a 1.5 °C warming occurs for the first time. Statistics are computed over all 26 climate models and all years within the given decade.

cumulative CO<sub>2</sub> budget estimates, which have larger remaining budgets than earlier estimates<sup>12,14</sup>. These, however, do not fundamentally change the need for strong near-term mitigation measures and technologies capable of enabling net-zero global CO<sub>2</sub> emissions near to mid-century if the considered emissions pathways are to be followed.

### Global and regional climate responses

Considering a subset of regions and extremes shown to retain particularly strong changes under a global warming of 1.5 °C or 2 °C<sup>4,37</sup>, Table 1 provides corresponding regional responses for the evaluated 1.5 °C- and 2 °C-compatible emissions pathways. Figures 2 and 3 display associated regional changes for a subset of considered extremes: temperature extremes (coldest nights in the Arctic, warmest days in the contiguous USA) and heavy precipitation (consecutive 5-day maximum precipitation in Southern Asia). Changes in hot extremes in central Brazil and in drought occurrence in the Mediterranean region are also provided in Table 1. We note that the spread displayed for single-scenario subsets in Figs. 2 and 3 corresponds to the spread of the global climate simulations of the 5th phase of the Coupled Model Intercomparison Project (CMIP5), underlying the derivation of the regional extremes for given global temperature levels<sup>4,37</sup> (see Box 1 for details).

In terms of the resulting global mean temperature increase, Fig. 2 shows that the difference between the 10% ‘worst-case’ and the 66% ‘probable’ outcomes of the scenarios is substantial, both for the 1.5 °C and 2 °C scenarios. Interestingly, the worst outcomes of the 1.5 °C scenarios are similar to the probable outcomes of the 2 °C scenarios. Indeed, both of these types of outcome show less than 2 °C warming by 2100, and approximately 2 °C warming in the overshoot phase, although the warming in the overshoot phase can be slightly higher for the worst-case 1.5 °C scenario than for the probable 2 °C scenarios assessed here. Hence, the scenarios aiming at limiting global warming to 1.5 °C also have a clear relevance for limiting global warming to 2 °C<sup>13</sup>, in that they ensure that the 2 °C threshold is not exceeded at the end of the twenty-first century. This contrasts with pathways designed to keep warming to 2 °C, but that have a 10% high-end (‘worst-case’) warming of more than 2.4 °C. This result is important when considering 2 °C warming as a ‘defence line’ that should not be exceeded<sup>2</sup>.

Assessing changes in regional extremes illustrates the importance of considering the geographical distribution of climate change in addition

to the global mean warming. Indeed, the average global warming does not convey the level of regional variability in climate responses<sup>4</sup>. By definition, because the global mean temperature is an average in time and space, there will be locations and time periods in which 1.5 °C warming is exceeded even if the global mean temperature rise is restrained to 1.5 °C. This is already the case today, at about 1 °C of global warming compared to the preindustrial period<sup>16</sup>. Similarly, some locations and time frames will display less warming than the global mean.

Extremes at regional scales can warm much more strongly than the global mean. For example, in scenarios compatible with 1.5 °C global warming, minimum night-time temperatures ( $T_{\text{night,min}}$ ) in the Arctic can increase by more than 7 °C at peak warming if the ‘probable’ (66th percentile) outcome of scenarios materializes, and more than 8 °C if the ‘worst-case’ (highest 10%, that is, 90th percentile) outcome of the scenarios materializes (Fig. 2). For the ‘worst-case’ outcome of scenarios considered to be compatible with warming of 2 °C, the changes in these cold extremes is even larger, and can reach more than 9 °C at peak warming (Fig. 2). Although the change is more limited for hot extremes (annual maximum mid-day temperature,  $T_{\text{day,max}}$ ) in the contiguous USA, it is nevertheless substantial. At peak warming, these hot extremes can increase by more than 4 °C for the probable 1.5 °C scenarios (the maximum in 66% of the cases), reaching 5 °C warming for the ‘worst-case’ 1.5 °C scenarios and slightly less for the highest ‘probable’ 2 °C scenarios. If the 10% ‘worst-case’ temperature outcome materializes after following a pathway that is considered compatible with 2 °C warming today, the temperature increase of the hottest days ( $T_{\text{day,max}}$ ) could exceed 5 °C at peak global warming in that region (Fig. 2).

These analyses also reveal the level of inter-model range in regional responses, when comparing the full spread of the CMIP5 distributions (Fig. 2). This interquartile range reaches about 2 °C for  $T_{\text{night,min}}$  in the Arctic and 1 °C for  $T_{\text{day,max}}$  in the contiguous USA at peak warming, that is, it is 2–4 times larger than the difference in global warming at 1.5 °C versus 2 °C. The intermodel range is also very large for changes in heavy precipitation in Southern Asia (Fig. 2), with an approximate doubling of the response at peak warming for the 75th quantile in the most sensitive models compared to the 25th quantile in the least sensitive models. This highlights that uncertainty in regional climate sensitivity to given global warming levels is an important component of uncertainty in impact projections in low-emissions scenarios (like uncertainty in mitigation pathways or the global transient climate response). Indeed, in cases

showing a high regional climate sensitivity (either owing to model specificities or internal climate variability), the tail values of the climate model distributions for 'probable' 1.5 °C-scenario outcomes overlap or even exceed likely values for the 'worst-case' 2 °C-scenario outcome (Fig. 2). This thus shows that even under the most stringent mitigation (1.5 °C) pathways, some risk of dangerous changes in regional extremes (that is, equivalent to or stronger than expected responses at 2 °C global warming) cannot be excluded.

While most climate change risk assessments factor in the inter-model range of regional climate responses, relatively few consider the effects of extreme weather, such as the temperature increase of the hottest days ( $T_{\text{day,max}}$ ). Recent literature highlights how these extreme events strongly influence levels of risk to human and natural systems, including crop yields<sup>38</sup> and biodiversity<sup>39</sup>, suggesting that the majority of risk assessments based on mean regional climate changes alone are conservative in that they do not incorporate the effects of extreme weather events. In addition, the co-occurrence of extreme events is also highly relevant for accurately assessing changes in risk, although analyses in this area are still lacking<sup>40,41</sup>.

Hence, the regional analyses of changes in extremes for scenarios aiming at limiting warming to 1.5 °C and 2 °C highlight the following main findings:

(1) Some regional responses of temperature extremes will be much larger than the changes in global mean temperature, by a factor of up to three ( $T_{\text{night,min}}$  in the Arctic).

(2) The regional responses at peak warming for scenarios that are today considered to be compatible with limiting warming to 1.5 °C (that is, having a 66% chance of stabilizing at 1.5 °C by 2100) can still involve an extremely large increase in temperature in some locations and time frames, in the worst case more than 8 °C for extreme cold night-time temperatures or up to 5 °C for daytime hot extremes (Fig. 2). We note that these numbers are substantially larger than present-day variability (see Supplementary Information).

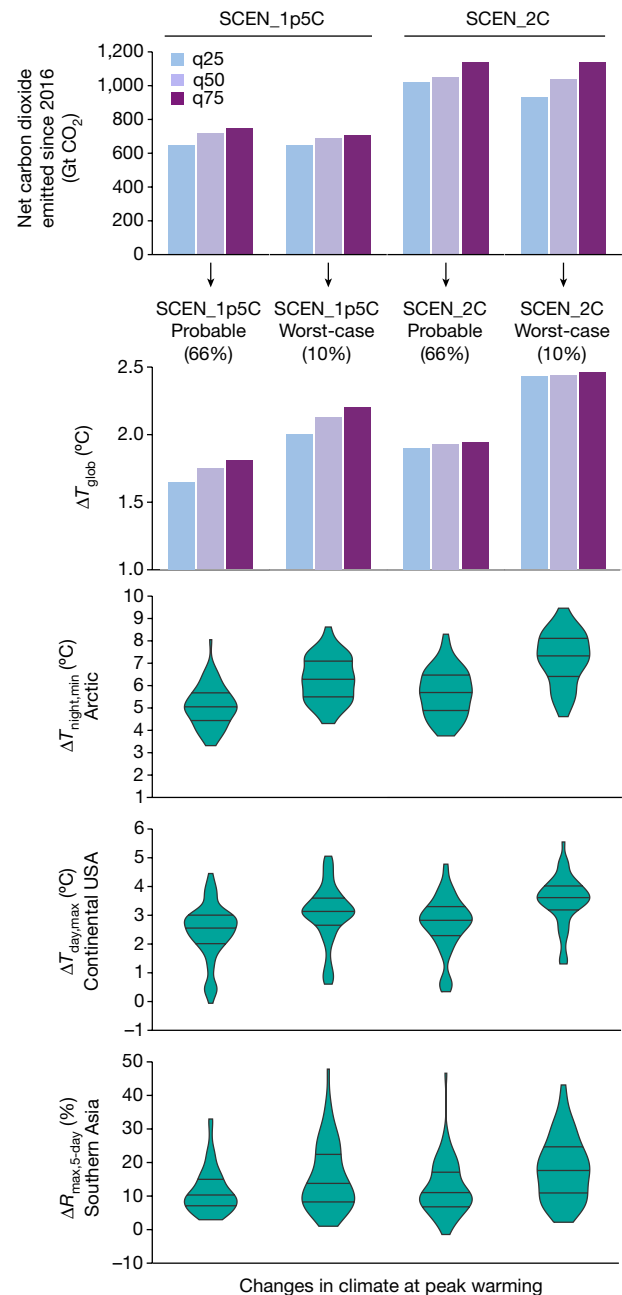
(3) The 10% highest-response ('worst-case') temperature outcome of pathways currently considered compatible with 1.5 °C warming is comparable with the 66th percentile ('probable') outcomes of scenarios that are considered compatible with limiting warming to below 2 °C, at global and regional scales. This indicates that pursuing a pathway compatible with warming of 1.5 °C can be considered a high-probability 2 °C pathway<sup>13</sup> that strongly increases the probability of avoiding the risks of a 2 °C warmer world.

### Realization at single locations and times

The analyses of Figs. 2 and 3 represent the statistical response over longer time frames. Several dominant patterns of response are documented in the literature<sup>4</sup>, for instance that land temperatures tend to warm more than global mean temperature on average, in particular with respect to hot extremes in transitional regions between dry and wet climates and with respect to coldest days at high latitudes (see also Figs. 2 and 3). Nonetheless, owing to internal climate variability (and in part model-based uncertainty), there may be large local departures from such a typical response at single points in time (any given year within a 10-year time frame), as displayed in Fig. 4. Many locations show a fairly large probability (25% chance) of temperature anomalies below 1.5 °C, and in some cases even smaller anomalies (mostly for the extreme indices). On the other hand, there is a similar probability (25%, or 75th percentile) that some locations can display temperature increases of more than 3 °C, and in some cases up to 7–9 °C for cold extremes. This illustrates that highly unusual and even unprecedented temperatures may occur even in a 1.5 °C climate. Although some of the patterns reflect what is expected from the median response<sup>4</sup>, the spread of responses is large in most regions.

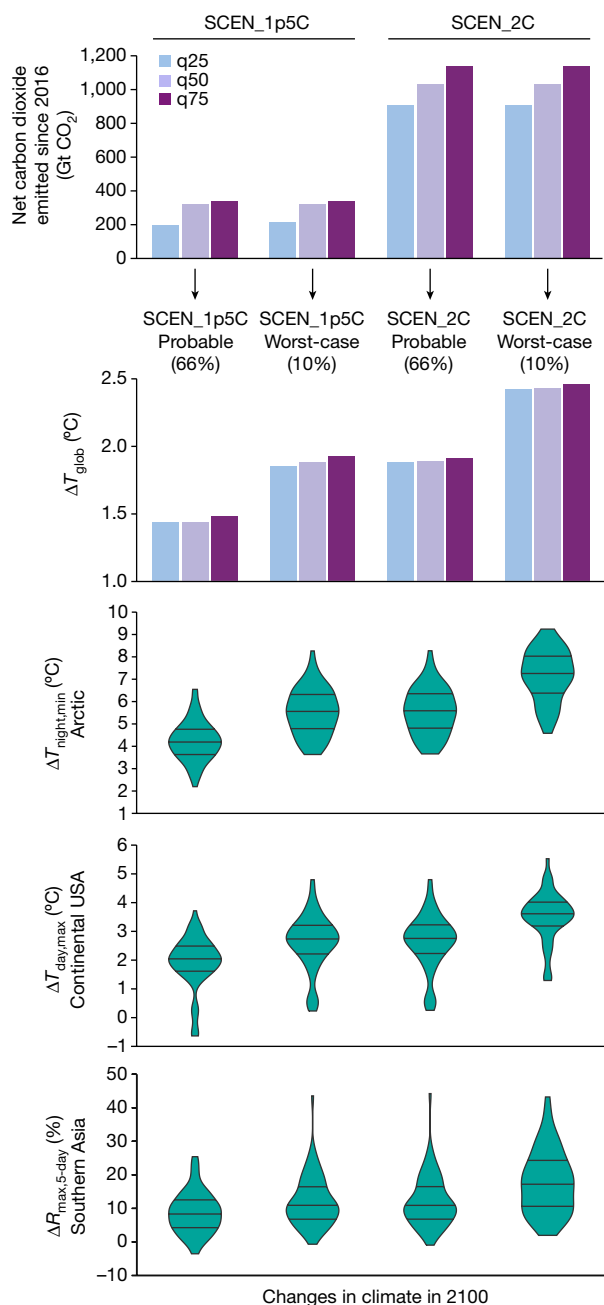
### Aspects insufficiently considered so far

The integrated assessment models used to derive the mitigation scenarios discussed here did not include several feedbacks that are present in the coupled human–Earth system. This includes, for example,



**Fig. 2 | Possible outcomes with respect to global temperature and regional climate anomalies from typical scenarios compatible with 1.5 °C warming and 2 °C warming at peak warming.** **a**, Net gigatonnes of CO<sub>2</sub> emitted until time of peak warming relative to 2016 (including carbon dioxide removal from the atmosphere) in scenarios from Table 1 (25th quantile (q25), median (q50), and 75th quantile (q75)). **b**, Global mean temperature anomaly at peak warming (q25, q50, q75). **c–e**, Regional climate anomalies at peak warming compared to the pre-industrial period corresponding to the median global warming of the second row (full range associated with different regional responses within CMIP5 multi-model ensemble displayed as violin plots; the median and interquartile ranges are indicated with horizontal dark grey lines). See Table 1 for more details.

biogeophysical impacts of land use<sup>26,27,28</sup>, potential competition for land between negative emission technologies and agriculture<sup>29,31</sup>, water availability constraints on energy infrastructure and bioenergy cropping<sup>30,31</sup>, regional implications of choices of specific scenarios for tropospheric aerosol concentrations, or behavioural and societal changes in anticipation of or in response to climate impacts<sup>33,42</sup>. For comprehensive assessments of the regional implications of mitigation and adaptation measures, such aspects of development pathways would need to be factored in.



**Fig. 3 | Possible outcomes with respect to global temperature and regional climate anomalies from typical scenarios compatible with 1.5°C warming and 2°C warming in 2100.** **a**, Net gigatonnes of CO<sub>2</sub> emitted by 2100 relative to 2016 (including carbon dioxide removal from the atmosphere) in scenarios from Table 1 (q25, q50, q75). **b**, Global mean temperature anomaly in 2100 (q25, q50, q75). **c–e**, Regional climate anomalies at peak warming compared to the pre-industrial period corresponding to the median global warming of the second row (full range associated with different regional responses within CMIP5 multi-model ensemble displayed as violin plots; the median and interquartile ranges are indicated with horizontal dark grey lines). See Table 1 for more details.

We note also that non-CO<sub>2</sub> greenhouse gas emissions need to be reduced jointly with CO<sub>2</sub>. The numbers in Table 1 consider budgets for cumulative CO<sub>2</sub> emissions taking into account consistent evolution of non-CO<sub>2</sub> greenhouse gas emissions. To compare the temperature outcome of pathways from many different forcings (such as methane and nitrous oxide), a CO<sub>2</sub>-only emission pathway that has the same radiative forcing can be found, which is termed ‘CO<sub>2</sub>-forcing equivalent emissions’<sup>43,44</sup>. Hence, stronger modulation in non-CO<sub>2</sub> greenhouse gas emissions could be considered in upcoming scenarios.

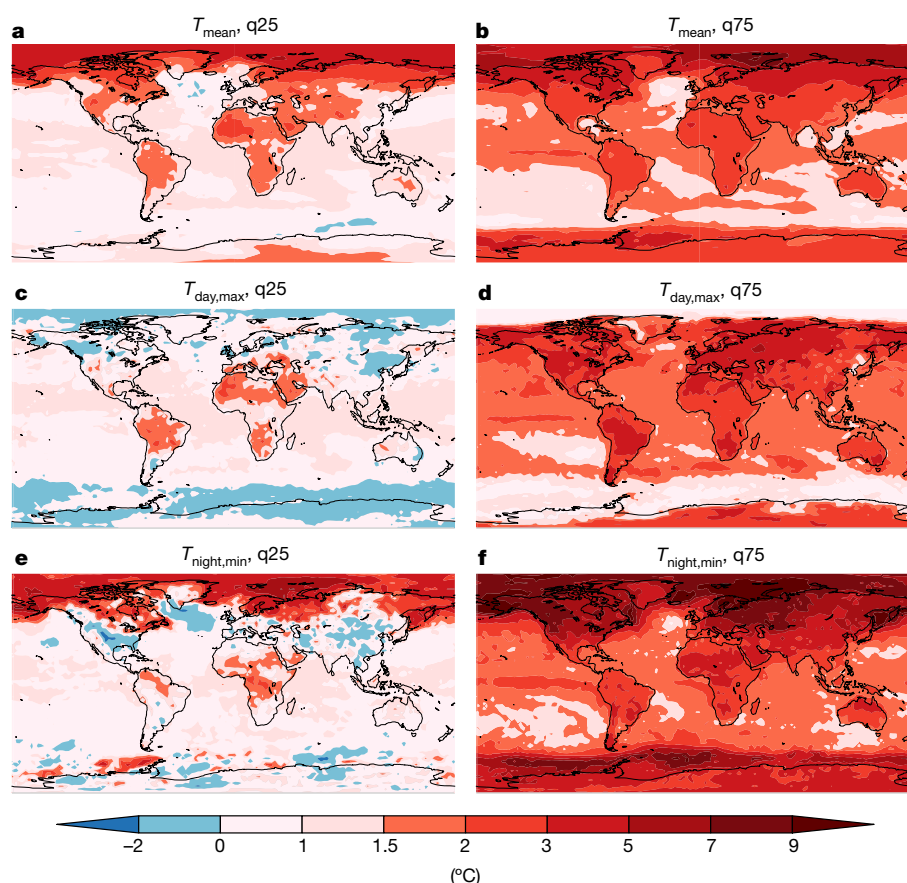
Furthermore, a continuous adjustment of mitigation responses based on the observed climate response (which can, for example, reduce present uncertainties regarding the global transient climate response) might be necessary to avoid undesired outcomes. Pursuing such ‘adaptive’ mitigation scenarios<sup>34</sup> would be facilitated by the global stocktake mechanism established in the Paris Agreement. Nonetheless, there are limits to possibilities for the adaptation of mitigation pathways, notably because some investments (for example, in infrastructure) are long-term, and also because the actual departure from a desirable pathway will need to be detected against the backdrop of internal climate variability. This variability can be large on decadal timescales—as illustrated by the recent so-called “hiatus” period<sup>45</sup>—but its effect on the assessment of mean global temperature anomalies can be minimized by using robust estimates of human-induced warming<sup>16</sup>. Hence, although adaptive mitigation pathways could provide some flexibility with which to avoid the highlighted ‘worst-case’ scenarios (Table 1), it is not yet clear to what extent they could be implemented in practice.

For a range of indicators, global mean temperature alone is not a sufficient indicator to describe climate impacts. CO<sub>2</sub>-sensitive systems, such as the terrestrial biosphere and agriculture systems, respond not only to the impact of warming but also to increased CO<sub>2</sub> concentrations. Although the potentially positive effects of CO<sub>2</sub> fertilization are not well constrained<sup>46</sup>, it appears that the impacts of anthropogenic emissions on those systems will depend not only on the warming inferred, but also on the CO<sub>2</sub> concentrations at which these warming levels are reached. Similarly, impacts on marine ecosystems depend on warming as well as on changes being driven by ocean acidification<sup>47</sup>.

Impacts on ocean and cryosphere will respond to warming with a substantial time lag. Consequently, ice sheet and glacier melting, ocean warming and as a result sea level rise will continue long after temperatures have peaked<sup>48</sup>. For some of these impacts, this may imply that the detectable effects of mitigation pathways may be limited in the short-term, but may turn out to be major effects in the long-term<sup>49</sup>. Large-scale oceanic systems will also continue to adjust over the coming centuries. One study identified a continued increase of extreme El Niño frequency in a peak-and-decline scenario<sup>50</sup>. The imprints on such time-lagged systems for different 1.5°C worlds are not well constrained at present.

### Assessing SRM

Compared to any mitigation options, climate interventions such as global SRM do not intend to reduce atmospheric CO<sub>2</sub> concentration itself but solely to limit global mean warming. Some studies<sup>51–53</sup> proposed that SRM may be used as a temporary measure to avoid global mean temperature exceeding 2°C. However, the use of SRM in the context of limiting temperature overshoot might create a new set of global and regional impacts, and could substantially modify regional precipitation patterns as compared to a world without SRM<sup>54,55</sup>. It would also have a high potential for cross-boundary conflicts because of positive, negative or undetectable effects on regional climate<sup>56</sup>, natural ecosystems<sup>57</sup> and human settlements. Hence, while the global mean temperature might be close to a 1.5°C warming under a given global SRM deployment, the regional implications could be very different from those of a 1.5°C global warming reached with early reductions of CO<sub>2</sub> emissions and stabilization of CO<sub>2</sub> concentrations. In some cases, some novel climate conditions would be created because of the addition of two climate forcings with different geographical footprints. Hence, a similar mean global warming may have very different regional implications (see Fig. 1b for an illustration) and in the case of SRM would be associated with substantial uncertainties in terms of regional impacts. Furthermore, SRM would not counter ocean acidification, which would continue unabated under enhanced CO<sub>2</sub> concentrations. Finally, there is also the issue that the sudden discontinuation of SRM measures would lead to a ‘termination problem’<sup>52,58</sup>, that is, a very rapid increase in global temperature and associated climate changes, which would have even greater impacts than a situation without SRM, owing to the rate of change. Together, this implies that the aggregated



**Fig. 4 | The stochastic noise and model-based uncertainty of realized climate at 1.5 °C.** Temperature with 25% chance of occurrence at any location within 10-year time frames corresponding to  $\Delta T_{\text{glob}} = 1.5$  °C (based on CMIP5 multi-model ensemble). The plots display at each location the 25th percentile (q25; **a**, **c**, **e**) and 75th percentile (q75; **b**, **d**, **f**)

environmental implications of an SRM world with 1.5 °C mean global temperature warming would probably be very different, and probably more detrimental and less predictable, from those of a 1.5 °C warmer world in which the global temperature is limited to 1.5 °C through decarbonization alone. Nonetheless, regional-scale changes in surface albedo may be worthwhile considering in order to reduce regional impacts in cities or agricultural areas<sup>21</sup>, although in-depth assessments on this topic are not yet available, and such modifications would be unlikely to affect global temperature substantially.

### Risks in 1.5 °C warmer worlds

1.5 °C warmer worlds will still present climate-related risks to natural, managed and human systems, as seen above. The magnitude of the overall risks and their geographical patterns in a 1.5 °C warmer world will, however, not only depend on the uncertainties in the regional climate that results from this level of warming. The magnitude of risk will also strongly depend on the approaches used to limit warming to 1.5 °C and on the wider context of societal development as it is pursued by individual communities and nations, and global society as a whole. Indeed, these can result in substantial differences in the magnitude and pattern of exposures and vulnerabilities<sup>59,60</sup>.

For natural ecosystems and agriculture, low-emissions scenarios can have a high reliance on land-use modifications (either for bioenergy production or afforestation<sup>25,29,61</sup>) that in turn can affect food production and prices through land-use competition effects<sup>29,31,62</sup>. The risks to human systems will depend on the ambition and effectiveness of implementing accompanying policies and measures that increase resilience to the risks of climate change and the potential trade-offs of mitigation. For example, large-scale deployment of BECCS could push Earth closer to the planetary boundaries for land-use change and

values of mean temperature ( $T_{\text{mean}}$ ; **a**, **b**), yearly maximum day-time temperature ( $T_{\text{day,max}}$ ; **c**, **d**), and yearly minimum night-time temperature ( $T_{\text{night,min}}$ ; **e**, **f**), sampled from all time frames with  $\Delta T_{\text{glob}} = 1.5$  °C in all Representative Concentration Pathway (RCP) 8.5 model simulations of the CMIP5 ensemble (see Box 1 for details).

freshwater, biosphere integrity and biogeochemical flows<sup>30</sup> (in addition to pressures associated with development goals<sup>63</sup>).

Also, the timing of when warming can be stabilized to 1.5 °C or 2 °C will influence exposure and vulnerability. For example, in a world pursuing a strong sustainable development trajectory, large increases in resilience by the end of the century would make the world less vulnerable overall<sup>59</sup>. Even under this pathway, rapidly reaching 1.5 °C would mean that some regions and sectors would require additional preparation to manage the hazards created by a changing climate.

### Commonalities of all 1.5 °C warmer worlds

Because human-caused warming linked to CO<sub>2</sub> emissions is now nearly irreversible for more than a thousand years<sup>64,65</sup>, the cumulative amount of CO<sub>2</sub> emissions is the prime determinant of long-lived permanent changes in the global mean temperature rise at Earth's surface. All 1.5 °C stabilization scenarios require net CO<sub>2</sub> emissions to be zero and non-CO<sub>2</sub> forcing to be capped to stable levels at some point<sup>64,66,67</sup>. This is also the case for stabilization scenarios at higher levels of warming (for example, at 2 °C); the only differences would be the time at which the net CO<sub>2</sub> budget becomes zero, and the cumulative CO<sub>2</sub> emissions emitted until that time. Hence, a transition to a decarbonization of energy use is necessary in all scenarios.

Article 4 of the Paris Agreement calls for net zero global greenhouse gas emissions to be achieved in the second half of the twenty-first century, which most plausibly requires some extent of negative CO<sub>2</sub> emissions to compensate for remaining non-CO<sub>2</sub> forcing<sup>13</sup>. The timing of when net zero global greenhouse gas emissions are achieved strongly determines the peak warming. All presently published scenarios compatible with 1.5 °C warming include carbon dioxide removal to achieve net-zero CO<sub>2</sub> emissions, to varying degrees. CO<sub>2</sub>-induced

warming by 2100 is determined by the difference between the total amount of CO<sub>2</sub> generated (which can be reduced by early decarbonization) and the total amount permanently stored out of the atmosphere, for example by geological sequestration. Current evidence indicates that at least some measure of carbon dioxide removal will be required to follow a emissions trajectory compatible with 1.5 °C warming.

### Towards a sustainable 1.5 °C warmer world

Emissions pathways limiting global warming to 1.5 °C allow us to avoid risks associated with higher levels of warming, but do not guarantee an absence of climate risks at the regional scale, and are also associated with their own set of risks with respect to the implementation of mitigation technologies, in particular related to land-use changes associated with for example, BECCS or competition for food production<sup>29–31,33</sup>.

Important aspects to consider when pursuing the goal of limiting warming to or below a global mean temperature level relate to how this goal is achieved and to the nature of emerging regional and sub-regional risks<sup>68–70</sup>. Also relevant are considerations of how the policies influence the resilience of human and natural systems, and which broader societal pathways are followed in terms of human development. Many but not all of these can be influenced directly through policy choices<sup>68–70</sup>. Internal climate variability as well as regional climate sensitivity, which display a substantial range between current climate models, are also important components of how risk will be realized. Explicitly illustrating the full range of possible outcomes of 1.5 °C warmer worlds is important for an adequate consideration of the implications of mitigation options by decision makers.

The time frame within which major mitigation measures need to be initiated varies with the scenario (Table 1). However, given the current state of knowledge about both the global and regional climate responses and the availability of mitigation measures, if the potential to limit warming to below 1.5 °C or 2 °C is to be maximized, emissions reductions in CO<sub>2</sub> and other greenhouse gases would need to start as soon as possible, leading to a global decline in emissions following 2020 at the latest. At the same time, if potential competition for land and water between negative emission technologies, agriculture and biodiversity conservation is to be avoided, mitigation would need to be carefully designed and regulated to minimize such competition, which could otherwise act to increase food prices and reduce ecosystem services (such as biodiversity, recreational uses and environmental functions). The remaining uncertainties underscore the need for continuous monitoring not just of global mean surface temperature, but also of the deployment and development of mitigation options, the resulting emissions reductions, and in particular of the intensity of global and regional climate responses and their sensitivity to climate forcing. Together with the choices made towards overall societal development, these various elements strongly co-determine the regional and sectoral magnitudes and patterns of risk at 2 °C and 1.5 °C of global warming.

### Code availability

The R code used to analyse MAGICC outputs in this paper is available from R.S. (roland.seferian@meteo.fr) on reasonable request. The scripts used for the regional analyses provided in Table 1 and Figs. 2 and 4 are available from R.W. (richard.wartenburger@env.ethz.ch) and S.I.S. (sonia.seneviratne@ethz.ch) upon request.

### Data availability

The data underlying the analyses of Table 1 and Figs. 2 and 3 are available on request. Emission data are available from the database accompanying ref. <sup>15</sup>, which presents pathways in line with 1.9 W m<sup>−2</sup> of radiative forcing in 2100, limiting warming to below 1.5 °C by 2100. Regional changes in climate extremes for different global warming levels derived following the methodology of refs <sup>4,37</sup> can be obtained from the associated database associated with the ERC DROUGHT-HEAT project (<http://www.drought-heat.ethz.ch>) and the software developed under ref. <sup>37</sup>.

Received: 1 November 2017; Accepted: 5 April 2018;  
Published online 6 June 2018.

1. United Nations Framework Convention on Climate Change (UNFCCC). *Adoption of the Paris Agreement FCCC/CP/2015/L.9/Rev.1* <http://unfccc.int/resource/docs/2015/cop21/eng/l09r01.pdf> (UNFCCC, 2015).
2. UNFCCC. *Report on the Structured Expert Dialogue on the 2013–2015 Review* <http://unfccc.int/resource/docs/2015/sb/eng/inf01.pdf> (UNFCCC, 2015).  
**This document prepared in advance of the Paris Agreement provides the underlying rationale for setting changes in global temperature as climate targets.**
3. Intergovernmental Panel on Climate Change (IPCC) in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Stocker, T. F. et al.) 3–29 (Cambridge Univ. Press, Cambridge, 2013).
4. Seneviratne, S. I., Donat, M. G., Pitman, A. J., Knutti, R. & Wilby, R. L. Allowable CO<sub>2</sub> emissions based on regional and impact-related climate targets. *Nature* **529**, 477–483 (2016).  
**This article highlights the large regional spread in climate model responses associated with given global temperature levels for specific regions and variables.**
5. Rogelj, J., Schleussner, C.-F. & Hare, W. Getting it right matters—temperature goal interpretations in geoscience research. *Geophys. Res. Lett.* **44**, 10662–10665 (2017).
6. Cowtan, K. & Way, R. G. Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends. *Q. J. R. Met. Soc.* **140**, 1935–1944 (2014).
7. Richardson, M., Cowtan, K., Hawkins, E. & Stolpe, M. B. Reconciled climate response estimates from climate models and the energy budget of Earth. *Nat. Clim. Chang.* **6**, 931–935 (2016).
8. Loarie, S. R. et al. The velocity of climate change. *Nature* **462**, 1052–1055 (2009).
9. LoPresti, A. et al. Rate and velocity of climate change caused by cumulative carbon emissions. *Environ. Res. Lett.* **10**, 095001 (2015).
10. Bowerman, N. H. A., Frame, D. J., Huntingford, C., Lowe, J. A. & Allen, M. R. Cumulative carbon emissions, emissions floors and short-term rates of warming: implications for policy. *Phil. Trans. R. Soc. A* **369**, 45–66 (2011).
11. Settele, J. et al. in *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Field, C. B. et al.) 271–359 (Cambridge Univ. Press, Cambridge, 2014).
12. Rogelj, J. et al. Energy system transformations for limiting end-of-century warming to below 1.5 °C. *Nat. Clim. Chang.* **5**, 519–527 (2015).
13. Schleussner, C.-F. et al. Science and policy characteristics of the Paris Agreement temperature goal. *Nat. Clim. Chang.* **6**, 827–835 (2016).  
**This article provides a discussion of the Paris Agreement from both scientific and policy perspectives.**
14. Clarke, L. et al. in *Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Edenhofer, O. et al.) 413–510 (Cambridge Univ. Press, Cambridge, 2014).  
**This chapter provides an overview of the scenarios considered compatible with limiting warming to 1.5 °C or 2 °C at the time of the IPCC AR5 report.**
15. Rogelj, J. et al. Scenarios towards limiting global mean temperature increase below 1.5 °C. *Nature Clim. Chang.* **8**, 325–332 (2018).  
**This article provides an overview on 1.5 °C scenarios from multiple models and under a wide range of socio-economic futures, revealing overall consistent results with previous publications<sup>12,14</sup> (see Box 1 and Supplementary Information).**
16. Haustein, K. et al. A real-time Global Warming Index. *Sci. Rep.* **7**, 15417 (2017).
17. Robinson, A., Calov, R. & Ganopolski, A. Multistability and critical thresholds of the Greenland ice sheet. *Nat. Clim. Chang.* **2**, 429–432 (2012).
18. Adger, W. N. et al. in *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Field, C. B. et al.) 755–791 (Cambridge Univ. Press, Cambridge, 2014).
19. Lawrence, D. M. et al. The Land Use Model Intercomparison Project (LUMIP) contribution to CMIP6: rationale and experimental design. *Geosci. Model Dev.* **9**, 2973–2998 (2016).
20. Pitman, A. J. et al. Uncertainties in climate responses to past land cover change: first results from the LUCID intercomparison study. *Geophys. Res. Lett.* **36**, L14814 (2009).
21. Seneviratne, S. I. et al. Land radiative management as contributor to regional-scale climate adaptation and mitigation. *Nat. Geosci.* **11**, 88–96 (2018).
22. Wang, Z. et al. Scenario dependence of future changes in climate extremes under 1.5 °C and 2 °C global warming. *Sci. Rep.* **7**, 46432 (2017).
23. Vogel, M. M. et al. Regional amplification of projected changes in extreme temperatures strongly controlled by soil moisture-temperature feedbacks. *Geophys. Res. Lett.* **44**, 1511–1519 (2017).
24. Deser, C., Knutti, R., Solomon, S. & Phillips, A. S. Communication of the role of natural variability in future North American climate. *Nat. Clim. Chang.* **2**, 775–779 (2012).
25. van Vuuren, D. P. et al. RCP2.6: exploring the possibility to keep global mean temperature increase below 2 °C. *Clim. Change* **109**, 95–116 (2011).
26. Hirsch, A. L., Wilhelm, M., Davin, E. L., Thiery, W. & Seneviratne, S. I. Can climate-effective land management reduce regional warming? *J. Geophys. Res. Atmos.* **122**, 2269–2288 (2017).

27. Hirsch, A. L. et al. Biogeophysical impacts of land-use change on climate extremes in low-emissions scenarios: results from HAPPI-Land. *Earth's Future* **6**, 396–409 (2018).
  28. Seneviratne, S. I. et al. Climate extremes, land-climate feedbacks, and land use forcing at 1.5°C. *Phil. Trans. R. Soc. A* **376**, <https://doi.org/10.1098/rsta.2016.0450> (2018).
  29. Smith, P. et al. Biophysical and economic limits to negative CO<sub>2</sub> emissions. *Nat. Clim. Chang.* **6**, 42–50 (2016).
  30. Heck, V., Gerten, D., Lucht, W. & Popp, A. Biomass-based negative emissions difficult to reconcile with planetary boundaries. *Nature Clim. Chang.* **8**, 151–155 (2018).
  31. Boysen, L. R. et al. The limits to global-warming mitigation by terrestrial carbon removal. *Earth's Future* **5**, 463–474 (2017).
  32. Obersteiner, M. et al. How to spend a dwindling greenhouse gas budget. *Nat. Clim. Chang.* **8**, 7–10 (2018).
  33. Van Vuuren, D. P. et al. Alternative pathways to the 1.5°C target reduce the need for negative emission technologies. *Nat. Clim. Chang.* **8**, 391–397 (2018).
  34. Millar, R. J. et al. Emission budgets and pathways consistent with limiting warming to 1.5°C. *Nat. Geosci.* **10**, 741–747 (2017).
  35. Matthews, H. D. et al. Estimating carbon budgets for ambitious climate targets. *Curr. Clim. Change Rep.* **3**, 69–77 (2017).
  36. Goodwin, P. et al. Pathways to 1.5°C and 2°C warming based on observational and geological constraints. *Nat. Geosci.* **11**, 102–107 (2018).
  37. Wartenburger, R. et al. Changes in regional climate extremes as a function of global mean temperature: an interactive plotting framework. *Geosci. Model Dev.* **10**, 3609–3634 (2017).
- This article is an extension of ref. 4, providing changes in a range of regional extremes as a function of global temperature changes based on simulations assessed in the IPCC AR5<sup>72</sup>.**
38. Deryng, D., Conway, D., Ramankutty, N., Price, J. & Warren, R. Global crop yield response to extreme heat stress under multiple climate change futures. *Environ. Res. Lett.* **9**, 034011 (2014).
  39. McDermott-Long, O. et al. Sensitivity of UK butterflies to local climatic extremes: which life stages are most at risk? *J. Anim. Ecol.* **86**, 108–116 (2017).
  40. AghaKouchak, A., Cheng, L., Mazdiyasni, O. & Farahmand, A. Global warming and changes in risk of concurrent climate extremes: insights from the 2014 California drought. *Geophys. Res. Lett.* **41**, 8847–8852 (2014).
  41. Zscheischler, J. & Seneviratne, S. I. Dependence of drivers affects risks associated with compound events. *Sci. Adv.* **3**, e1700263 (2017).
  42. Beckage, B. et al. Linking models of human behaviour and climate alters projected climate change. *Nature Clim. Chang.* **8**, 79–84 (2018).
  43. Jenkins, S., Millar, R. J., Leach, N. & Allen, M. R. Framing climate goals in terms of cumulative CO<sub>2</sub>-forcing-equivalent emissions. *Geophys. Res. Lett.* **45**, 2795–2804 (2018).
  44. Fuglestad, J. et al. Implications of possible interpretations of “greenhouse gas balance” in the Paris Agreement. *Phil. Trans. R. Soc. A* **376**, <https://doi.org/10.1098/rsta.2016.0445> (2018).
  45. Medhaug, I., Stolpe, M. B., Fischer, E. M. & Knutti, R. Reconciling controversies about the ‘global warming hiatus’. *Nature* **545**, 41–47 (2017).
  46. Smith, K. et al. Large divergence of satellite and Earth system model estimates of global terrestrial CO<sub>2</sub> fertilization. *Nature Clim. Chang.* **6**, 306–310 (2016).
  47. Gattuso, J.-P. et al. Contrasting futures for ocean and society from different anthropogenic CO<sub>2</sub> emissions scenarios. *Science* **349**, aac4722 (2015).
  48. Clark, P. U. et al. Consequences of twenty-first-century policy for multi-millennial climate and sea-level change. *Nat. Clim. Chang.* **6**, 360–369 (2016).
  49. Marzeion, B., Kaser, G., Maussion, F. & Champollion, N. Limited influence of climate change mitigation on short-term glacier mass loss. *Nature Clim. Chang.* **8**, 305–308 (2018).
  50. Wang, G. et al. Continued increase of extreme El Niño frequency long after 1.5°C warming stabilization. *Nature Clim. Chang.* **7**, 568–572 (2017).
  51. Boucher, O., Lowe, J. A. & Jones, C. D. Implications of delayed actions in addressing carbon dioxide emission reduction in the context of geo-engineering. *Clim. Change* **92**, 261–273 (2009).
  52. Keith, D. W. & MacMartin, D. G. A temporary, moderate and responsive scenario for solar geoengineering. *Nature Clim. Chang.* **5**, 201–206 (2015).
  53. Tilmes, S., Sanderson, B. M. & O'Neill, B. C. Climate impacts of geoengineering in a delayed mitigation scenario. *Geophys. Res. Lett.* **43**, 8222–8229 (2016).
  54. Ferraro, A. J. & Griffiths, H. G. Quantifying the temperature-independent effect of stratospheric aerosol geoengineering on global-mean precipitation in a multi-model ensemble. *Environ. Res. Lett.* **11**, 34012 (2016).
  55. Davis, N. A., Seidel, D. J., Birner, T., Davis, S. M. & Tilmes, S. Changes in the width of the tropical belt due to simple radiative forcing changes in the GeoMIP simulations. *Atmos. Chem. Phys.* **16**, 10083–10095 (2016).
  56. Lo, Y. T. E., Charlton-Perez, A. J., Lott, F. C. & Highwood, E. J. Detecting sulphate aerosol geoengineering with different methods. *Sci. Rep.* **6**, 39169 (2016).
  57. Muri, H., Kristjánsson, J. E., Storelvmo, T. & Pfeffer, M. A. The climatic effects of modifying cirrus clouds in a climate engineering framework. *J. Geophys. Res.* **119**, 4174–4191 (2014).
  58. Trisos, C. H. et al. Potentially dangerous consequences for biodiversity of solar geoengineering implementation and termination. *Nat. Ecol. Evol.* **2**, 475–482 (2018).
  59. O'Neill, B. C. et al. The roads ahead: narratives for shared socioeconomic pathways describing world futures in the 21st century. *Glob. Environ. Change* **42**, 169–180 (2017).
  60. Byers, E. A. et al. Global exposure and vulnerability to multi-sector development and climate change hotspots. *Environ. Res. Lett.* **13**, 055012 (2018).
  61. Popp, A. et al. Land-use futures in the shared socio-economic pathways. *Glob. Environ. Change* **42**, 331–345 (2017).
  62. Muratori, M., Calvin, K., Wise, M., Kyle, P. & Edmonds, J. Global economic consequences of deploying bioenergy with carbon capture and storage (BECCS). *Environ. Res. Lett.* **11**, 95004 (2016).
  63. O'Neill, D. W., Fanning, A. L., Lamb, W. F. & Steinberger, J. K. A good life for all within planetary boundaries. *Nat. Sustain.* **1**, 88–95 (2018).
  64. Matthews, H. D. & Caldeira, K. Stabilizing climate requires near-zero emissions. *Geophys. Res. Lett.* **35**, L04705 (2008).
  65. Solomon, S., Plattner, G.-K., Knutti, R. & Friedlingstein, P. Irreversible climate change due to carbon dioxide emissions. *Proc. Natl Acad. Sci. USA* **106**, 1704–1709 (2009).
  66. Matthews, H. D., Gillett, N. P., Stott, P. A. & Zickfeld, K. The proportionality of global warming to cumulative carbon emissions. *Nature* **459**, 829–832 (2009).
  67. Allen, M. R. et al. Warming caused by cumulative carbon emissions towards the trillionth tonne. *Nature* **458**, 1163–1166 (2009).
  68. Denton, F. et al. in *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Field, C. B. et al.) 1101–1131 (Cambridge Univ. Press, Cambridge, 2014).
  69. Fleurbay, M. et al. in *Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Edenhofer, O. et al.) 283–350 (Cambridge Univ. Press, Cambridge, 2014).
  70. O'Brien, K. et al. in *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation* (eds Field, C. B. et al.) A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change (IPCC) 437–486 (Cambridge Univ. Press, Cambridge, 2012).
  71. Meinshausen, M., Raper, S. C. B. & Wigley, T. M. L. Emulating coupled atmosphere-ocean and carbon cycle models with a simpler model, MAGICC6—part 1: model description and calibration. *Atmos. Chem. Phys.* **11**, 1417–1456 (2011).
  72. IPCC. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Stocker, T. F. et al.) 1–1535 (Cambridge Univ. Press, Cambridge, 2013).
  73. Le Quéré, C. et al. Global carbon budget 2016. *Earth Syst. Sci. Data* **8**, 605–649 (2016).
  74. Keenan, T. F. et al. Recent pause in the growth rate of atmospheric CO<sub>2</sub> due to enhanced terrestrial carbon uptake. *Nat. Commun.* **7**, 13428 (2016).

**Acknowledgements** S.I.S. and R.W. acknowledge the European Research Council (ERC) ‘DROUGHT-HEAT’ project funded by the European Community’s Seventh Framework Programme (grant agreement FP7-IDEAS-ERC-617518). J.R. acknowledges the Oxford Martin School Visiting Fellowship programme for support. R.S. acknowledges the European Union’s H2020 project CRESCENDO “Coordinated Research in Earth Systems and Climate: Experiments, Knowledge, Dissemination and Outreach” (grant agreement H2020-641816). O.H.G. acknowledges support of the Australia Research Council Laureate program. This work contributes to the World Climate Research Programme (WCRP) Grand Challenge on Extremes. We acknowledge the WCRP Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modelling groups for producing and making available their model output. For CMIP the US Department of Energy’s Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals.

**Reviewer information** Nature thanks S. Davis, K. Tachiiri and the other anonymous reviewer(s) for their contribution to the peer review of this work.

**Author contributions** S.I.S. coordinated the design and writing of the article, with contributions from all co-authors. J.R. provided the emissions scenario data processed in Table 1. R.S. computed the scenario summary statistics of Table 1. R.W. computed the regional projections statistics of Table 1, as well as Figs. 2–4. S.I.S. prepared Fig. 1, with support from P.T. and J.R. J.R., R.S., M.A., M.C. and R.M. co-designed the analyses of emissions scenarios. K.L.E., N.E., O.H.G., A.J.P., C.F.S., P.T. and R.F.W. provided assessments on physical, ecosystem and human impacts. S.I.S. drafted the first version of the manuscript, with inputs from J.R., R.S. and M.A. All authors contributed to and commented on the manuscript.

**Competing interests** The authors declare no competing interests.

#### Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41586-018-0181-4>.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>.

**Correspondence and requests for materials** should be addressed to S.I.S.

**Publisher’s note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

# Long-distance navigation and magnetoreception in migratory animals

Henrik Mouritsen<sup>1,2\*</sup>

**For centuries, humans have been fascinated by how migratory animals find their way over thousands of kilometres. Here, I review the mechanisms used in animal orientation and navigation with a particular focus on long-distance migrants and magnetoreception. I contend that any long-distance navigational task consists of three phases and that no single cue or mechanism will enable animals to navigate with pinpoint accuracy over thousands of kilometres. Multiscale and multisensory cue integration in the brain is needed. I conclude by raising twenty important mechanistic questions related to long-distance animal navigation that should be solved over the next twenty years.**

Each year, billions of small songbirds (Fig. 1a), with ‘birdbrains’ weighing only a few grams, leave their Arctic and temperate breeding areas to overwinter in the tropics and subtropics. Most migrate at night, and young birds do so without regular contact with experienced individuals. Thus, their navigational capabilities must be innate or learned before their first departure<sup>1–5</sup>. After having completed one round trip, many adult birds are able to navigate with an ultimate precision of centimetres over distances of 5,000 km or more<sup>6</sup>. Other impressive navigational tasks mastered by birds include bar-tailed godwits (*Limosa lapponica*, Fig. 1b) migrating from Alaska to New Zealand in a single non-stop flight lasting 7–9 days and nights<sup>7</sup>, arctic terns (*Sterna paradisaea*) breeding around the North Pole and wintering around the South Pole<sup>8</sup>, and seabirds (Fig. 1c) flying more than 100,000 km per year to return to tiny islands in the middle of vast oceans to breed<sup>9,10</sup>.

Even insects with much simpler brains than birds are capable of performing impressive navigational tasks<sup>11–18</sup>. In autumn, Monarch butterflies (*Danaus plexippus*, Fig. 1d) migrate from the USA and Canada to very specific overwintering trees in Mexico, up to 3,000 km away<sup>11</sup>. A year later, the third-to-fifth-generation descendants of the previous year’s autumn migrants return to the exact same trees in Mexico<sup>11</sup>. A similarly impressive return migration—but involving only a single generation—occurs in Southeast Australia, where millions of Bogong moths (*Agrotis infusa*, Fig. 1e) fill the night skies on their way to and from their yearly aestivation caves in the Snowy Mountains<sup>18</sup>. Recently, Chapman et al.<sup>17,19,20</sup> demonstrated that directed long-distance return migrations are also widespread among high-flying insects. These movements of trillions of individual insects are critical for understanding both natural and man-made ecosystems<sup>21</sup>.

In the ocean, Salmonid fish (Fig. 1f) and sea turtles (Fig. 1g), for instance, return to their natal streams or beaches over thousands of kilometres<sup>22–25</sup> and many dispersing coral reef fish larvae relocate their natal reefs after being at the mercy of sea currents for weeks<sup>26–28</sup>.

To complete their long voyages, migratory animals have developed elaborate abilities to detect a variety of sensory cues, to integrate these signals within their nervous systems, and to use them as part of highly efficient navigational strategies<sup>1,3,4,10,17,29–32</sup>. Navigation skills are also vitally important to non-migratory animals of almost any class<sup>1,13,14,33,34</sup>. However, this review focuses primarily on long-distance navigation and homing. After discussion of the basic principles underlying these processes, I discuss how animals use, detect and process the main types of

navigation-relevant cue. I consider magnetic cues in more detail than other cues because the sensory mechanisms that underlie sight, olfaction and hearing are generally understood. By contrast, even though a lot of progress has been made recently, the mechanisms by which animals sense the geomagnetic field remains one of the most fundamentally important questions in sensory biology. I also highlight twenty of the most important outstanding mechanistic questions that remain to be answered (Box 1; denoted as ‘question 1’ and so on throughout the Review).

## Studying navigation

### Navigation and orientation

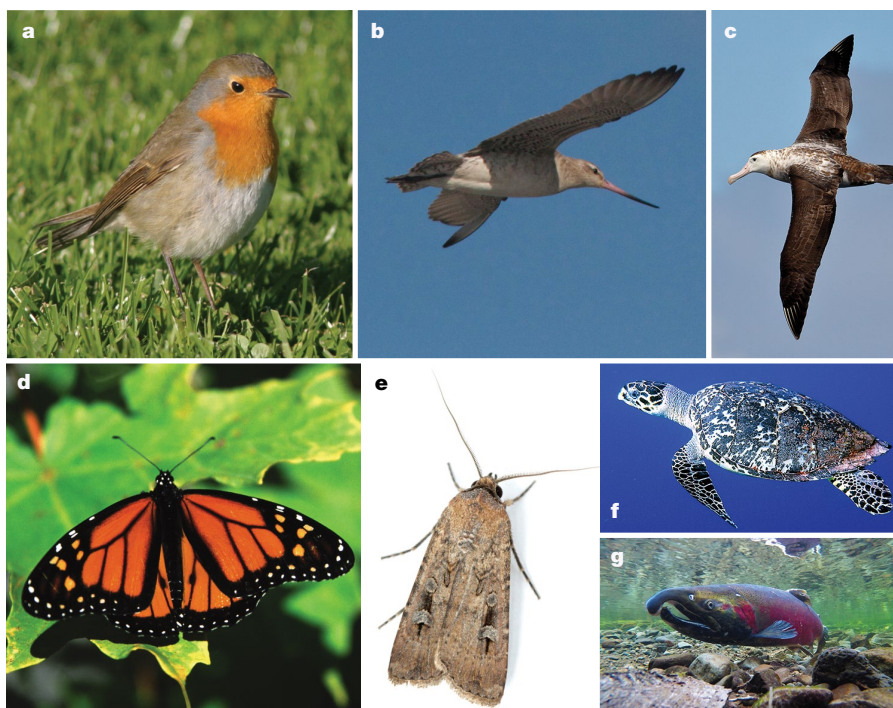
The terms ‘navigation’ and ‘orientation’ are used inconsistently in different fields. Here, ‘orientation’ means that only the direction of movement is being determined. To perform ‘true navigation’, animals need first to determine their location (map position) and then the compass direction to their goal<sup>3,31,35</sup>. True navigators can correct for displacements during any phase of their journey<sup>3,31,35–37</sup>. ‘Navigation’ is used for anything within the continuum between true navigation and pure compass orientation.

### Maps and compasses

Map and compass information are often determined independently<sup>1,3,4,31</sup>. To get a sense of direction, only a reference compass direction, such as magnetic and/or geographical North, needs to be determined, which an animal can then use to orient in any desired direction. Location can be determined in various ways. In some animals, location is defined relative to home<sup>1,13,14,34</sup>, whereas many experienced migrants have developed large-scale, probably multisensory and multicoordinate maps, which can be extrapolated to correct for displacements, even at unfamiliar locations<sup>1,3,4,10,30,31,36–38</sup>.

For instance, the angle of the celestial rotation centre above the horizon, geomagnetic field intensity, and geomagnetic inclination angle all gradually increase from south to north in most parts of the world<sup>11,3,30,38</sup>. Thus, higher or lower values indicate displacement to the north or south, respectively. How long-distance migrants determine longitude (east–west position) is much less clear (question 19). Magnetic declination is an excellent east–west cue in some parts of the world, and experienced Eurasian reed warblers seem to use magnetic declination as part of their map<sup>38</sup>. Because magnetic declination is the angular deviation between magnetic and geographical North, map and compass cues

<sup>1</sup>Institut für Biologie und Umweltwissenschaften, Universität Oldenburg, Oldenburg, Germany. <sup>2</sup>Research Centre for Neurosensory Sciences, University of Oldenburg, Oldenburg, Germany. e-mail: henrik.mouritsen@uol.de



**Fig. 1 | Some of the world's most famous long-distance navigators.** **a**, European robin (*Erithacus rubecula*). **b**, Bar-tailed godwit (*Limosa lapponica*). **c**, Wandering albatross (*Diomedea exulans*). **d**, Monarch butterfly (*Danaus plexippus*). **e**, Bogong moth (*Agrotis infusa*). **f**, Sea turtle

(*Eretmochelys imbricata*). **g**, Salmon (*Oncorhynchus kisutch*). Photographs by H.M. (**a**, **b**, **d**); E. Dunens (**c**); A. Narendra (**e**); Adam (**f**); and the Bureau of Land Management Oregon and Washington (**g**). (**c**, **f**, **g**: <https://creativecommons.org/licenses/by/2.0/>).

might not always be as separable as previously thought. Experimental compass manipulations could also influence the map.

### Experienced versus naïve animals

When studying long-distance navigation, it is important to consider whether animals are travelling for the first time. Animals such as migratory insects and coral reef fish larvae are always inexperienced migrants, as they complete only a single return journey or less<sup>11,18–20,39</sup>. By contrast, most migratory birds and sea turtles make several similar journeys<sup>1,3,4,30,31</sup>.

First-time migrants must use relatively simple orientation systems based on information inherited or learned before departure. Young night-migratory songbirds inherit their migratory direction and distance<sup>2</sup>, but the genes underpinning this have not been identified<sup>40</sup> (question 16). Inexperienced migrants cannot have a detailed map of their migration route, but could have inherited simple cue values for the goal and/or a few 'signposts' and associated these with adaptive behaviours, such as the responses of hatchling sea turtles to magnetic parameters<sup>22,30,41</sup>. Inexperienced bird migrants usually follow experienced companions or rely on a simple clock-and-compass strategy (vector navigation) using only an innate circannual clock and compass orientation programmes, but no map. They are therefore, except for a few emergency plans, unable to correct for geographical displacement<sup>3,4,31,36,38,42–45</sup>. It remains unclear exactly which combination of sensory parameters triggers the start and stop of the first natural migration (question 17).

By contrast, many experienced migrants travelling for the second or later time have experienced cue gradients and generated a map that they can use to correct even for displacements to unknown locations. They can thus perform true navigation<sup>1,3,4,31,36–38,43,45–48</sup>.

### The three phases of a navigational task

Navigational cues that can be used over thousands of kilometres differ from those that are useful over a few kilometres, metres, or centimetres over time-scales of a few seconds, minutes or hours (Table 1). Furthermore, animals mostly use quite simple navigational strategies

that are good enough to solve the tasks needed for survival, but not 'perfect' mathematical solutions. Consequently, a succession of at least three different phases or stages is needed to account for the pinpoint accuracy of experienced long-distance migrants<sup>32</sup> (Fig. 2). The three phases are: (1) a long-distance phase; (2) a narrowing-in or homing phase; and (3) a pinpointing-the-goal phase. To achieve a holistic understanding of animal navigation, all phases need to be understood, and a comparative approach is needed to evaluate whether species, groups or classes of animals use similar or different solutions.

The long-distance phase refers to navigation far away from the animal's home ranges and it usually relies on global or regionally stable cues such as celestial and/or geomagnetic information. Simple, compass-based, vector orientation relying on an inherited initial direction<sup>2,3,11,27,31</sup> seems to be the only mechanism available to many inexperienced animals that travel without experienced companions<sup>2,3,27,31,36,39,42–44</sup>. By contrast, experienced animals can often modify their compass headings on the basis of learned map information<sup>3,4,31,36–38,43,47–49</sup>. During the narrowing-in or homing phase, in or near a familiar home range, learned local gradient maps that rely on a variety of senses and environmental cues are usually important<sup>1,10,32,34,39,49</sup>. The pinpointing-the-goal phase is mostly based on remembering very specific visual landmarks and/or the odours of a specific location<sup>25,50,51</sup>.

The three navigational phases seem quite universal. Night-migratory songbirds use mainly celestial and magnetic cues during the long-distance phase<sup>1,3,4,31,37,38</sup>, a variety of learned, multisensory, local gradient maps during their homing phase<sup>1,32,49</sup>, and visual landmarks to find their nest or sleeping perch during the pinpointing-the-goal phase<sup>32</sup>.

Monarch butterflies use a time-compensated sun compass during the long-distance phase<sup>12,52,53</sup>. Monarchs do not like to cross large bodies of water (the Gulf of Mexico constrains movement towards the southeast) or to fly over high mountains (the Rocky Mountains limits them to the west). The resulting geographic funnelling effect brings the monarchs to within a couple of hundred kilometres of their wintering range<sup>39</sup>. How the later parts of the narrowing-in and

**Table 1 | Examples of typical cues that are relevant during the three phases of a long-distance navigational task.**

	Magnetic	Visual	Olfactory	Other
<b>Long-distance phase</b>	Horizontal direction; inclination angle; intensity; declination	Celestial cues related to the stars and sun; coastlines and major mountain ranges as physical constraints associated with simple response rules	Probably not useful far from familiar locations and on first time migratory journeys, probably useful at or near familiar routes	
<b>Narrowing-in or homing phase</b>	Horizontal direction; inclination and intensity down to a scale larger than 10–50 km; strong magnetic anomalies	Celestial cues; familiar leading lines (rivers, mountain ranges, coast lines, forest borders, roads, and so on); familiar beacons (specific forests, hills, lakes, buildings, and so on)	Natural olfactory gradients; olfactory 'landscapes'	Water depth; salinity; regional sound cues
<b>Pinpointing-the-goal phase</b>	Probably not useful as map cues on this scale	Very local familiar landmarks (for example, a specific tree, branch, or nest hole; a cave entrance; a small hill; a specific coral)	Local odours (for example, of home habitats or conspecifics)	Local sound cues; microclimate; waves; tidal flows

If a long-distance navigational task is split into several legs with specific intermediate goals, the three phases could be repeated several times before reaching the final goal. For references, see main text.

pinpointing-the-goal phases work in these one-time migrants is currently unknown (question 20). The latter could be based on a combination of attraction to smells left by previous generations of conspecifics beaconing from the wintering trees and searching for the right microclimate and tree species.

Salmon might use an innate signpost 'map' coupled with adaptive compass responses similar to those of sea turtles<sup>22</sup> to stay within a suitable oceanic range and to return as adults to the approximate location of the river mouth<sup>23</sup>. At this point, their navigational strategy changes to one based mainly on chemical or olfactory cues, which they use to home in on the exact spawning ground where they were born<sup>25,54,55</sup>. Coral reef fish larvae first seem to use an innate celestial and magnetic compass direction to relocate the vicinity of the reef<sup>27,28</sup>, then olfactory<sup>26</sup> and/or auditory cues<sup>56</sup> to narrow in on the reef, and finally vision to locate a suitable microhabitat within the reef.

In summary, several cues are often used together during a phase, and the cues, brain-processing strategies, and behaviours involved vary substantially between phases in most cases. What determines when an animal switches from one navigational phase to the next, and how processing strategies in the nervous system transition between phases, remain exciting open questions (questions 10, 11, 18).

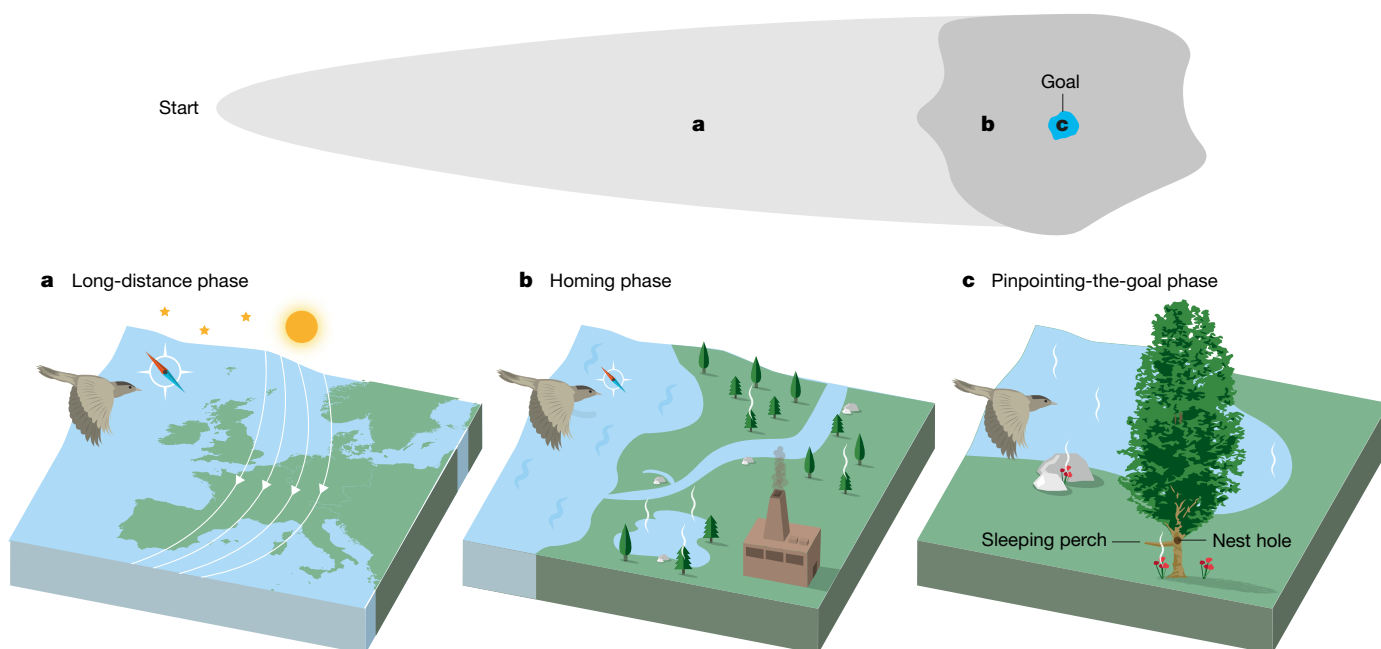
Owing to the three navigational phases, it is extraordinarily unlikely that a single sense or cue is used exclusively throughout a journey. One

consequence of this is that animals tested at the wrong location relative to where the relevant phase takes place in nature may not reveal their true abilities during that phase. Testing of animals during different phases or at wrong locations might explain some of the apparent contradictions in the long-distance navigation literature.

### Magnetic cues and how they are sensed

The Earth's magnetic field, also called the geomagnetic field, is shaped as if a big bar magnet were placed at the centre of the Earth<sup>4,57</sup>. The geomagnetic field provides omnipresent information, which can help animals to navigate. Magnetic direction (polarity) and/or inclination angle (the angle between the field lines and the Earth's surface) can be used to determine a favourable direction of movement<sup>1,4,57,58</sup>. Total magnetic intensity, inclination angle, and magnetic declination can help animals to determine position<sup>1,4,22,30,38,41,48,57</sup>.

Birds<sup>1,58,59</sup>, sea turtles<sup>60</sup>, fish<sup>28</sup> and amphibians<sup>61</sup> can use magnetic polarity and/or inclination angle as a reference direction for a magnetic compass<sup>1,57</sup>. Likewise, birds<sup>1,48,62</sup>, sea turtles<sup>22,24,30,41</sup>, fish<sup>23</sup> and amphibians<sup>46</sup> can use magnetic parameters to determine their position. By contrast, it is less clear whether long-distance migratory insects can use magnetic compass and/or map cues<sup>39,52,63</sup>. As the geomagnetic field, on average, varies only by approximately 3 nT km<sup>-1</sup> and 0.009° km<sup>-1</sup> on the north–south axis and much less east–west, and owing to regular stochastic



**Fig. 2 | The three different phases of a long-distance navigational task and examples of the typical cues used.** **a**, During the long-distance phase, celestial and magnetic compass and map cues are very important and landmarks such as coastlines can function as physical constraints. **b**, During the homing phase, compasses are usually still important and

regional map cues such as olfactory and visual landmarks, olfactory gradients, strong magnetic anomalies, and soundscapes become important. **c**, During the pinpointing-the-goal phase, specific within-habitat cues such as a cave entrance, a specific tree, or a smelly lake are needed to locate, for example, a nest hole or sleeping perch.

variations in the geomagnetic field of 30–100 nT in variable directions, it is hard to understand how a magnetic map could have an accuracy better than 10–30 km in fast-moving animals<sup>4,57,64</sup>. Some newts and pigeons seem to be able to use a magnetic map over shorter distances<sup>46,62</sup>. Unless the magnetic gradients are locally very steep and/or slowly moving animals improve resolution by averaging over many measurements, it remains to be understood how the magnetic spatial signal can be distinguished from temporal variability on a scale of less than 10–30 km (question 7). Thus, magnetic maps seem to be primarily relevant for the long-distance and/or far-distance homing phases, at least for fast-moving animals. Finally, a vast number of organisms, from magnetotactic bacteria<sup>65</sup> to mammals<sup>66</sup>, align themselves with the magnetic field. Thus, many animals can detect and use the geomagnetic field for orientation and navigation, but how do they detect magnetic field parameters?

The geomagnetic field penetrates biological materials. Consequently, the primary sensors could be located anywhere inside an animal's body. Considering the anatomical constraints and known structures found within small animals, it is not obvious how biological materials can reliably detect the 25,000–65,000 nT geomagnetic field (questions 1–8) in the presence of thermal fluctuations (energy  $\approx k_B T$  (Boltzmann's constant multiplied by the temperature in degrees Kelvin)) and other sources of noise<sup>67–72</sup>. Only three mechanisms are currently considered to be physically viable: (1) induced electrical fields detected by highly sensitive electroreceptors; (2) magnetic-particle-based magnetoreception; and (3) radical-pair-based magnetoreception.

### Electromagnetic induction

Electromagnetic induction is the production of voltage across an electrical conductor moving through a static magnetic field. A 'biological wire' occurs in elasmobranch fish (sharks, skates, and rays) in which highly conductive pores connect the electrosensitive ampullae of Lorenzini with seawater, which acts as reference potential (ground) against which induced voltages in the pores can be measured. However, it is not known whether these structures are used as magnetoreceptors<sup>73</sup>. Here is a potentially exciting research area ready for someone to take a closer look at using modern methods (question 5). It is difficult to imagine how non-aquatic animals could use induction to sense the geomagnetic field. As air has low conductivity, large internal ring-shaped structures filled with conductive liquid would be needed<sup>74</sup>, but no such structures have been reported. Thus, for terrestrial animals another mechanism must be responsible for magnetoreception<sup>57,67</sup>.

### Magnetic-particle-based magnetoreception

The discovery of magnetotactic bacteria, which build intracellular chains of magnetite ( $\text{Fe}_3\text{O}_4$ ) particles (magnetosomes), demonstrates that organisms can synthesize magnetic crystals that could act as compass needles<sup>65,75</sup>. Since the discovery of magnetotactic bacteria, magnetite and/or iron oxides have been detected in almost every animal carefully investigated<sup>76,77</sup>. However, the mere presence of iron oxides, or even magnetite, does not indicate that such particles are relevant for magnetoreception<sup>57,71,78–81</sup>. Iron homeostasis is important for organism function and iron oxides may just be a way for organisms to deposit excess iron<sup>57,78</sup>. Only if magnetic particles are located inside cells at consistent and specific locations in many individuals of the same species and are associated with the nervous system (question 6) can the particles qualify as serious magnetosensory candidates<sup>57,71,78,79,81</sup>.

Currently, the most promising magnetic-particle-based magnetoreceptor candidate structures are those described in the olfactory epithelium of fish<sup>80,82</sup> (but see<sup>81</sup>). Iron-rich structures associated with the ophthalmic branch of the trigeminal nerve in birds were also thought to be magnetoreceptors<sup>83</sup>. However, recent findings suggest that these structures are associated not with neurons but merely with macrophages<sup>78</sup>. It has also been suggested that the avian lagena (a part of the bird vestibular system) plays a role in magnetoreception<sup>84</sup>. Because mole rats, fish and sea turtles seem to use a magnetic polarity compass in complete darkness, it is most likely that they use magnetic-particle-based magnetoreception<sup>60,82,85,86</sup>. If magnetic sensory particles

exist in higher animals, the magnetic signal would be expected to be transduced by opening or closing mechanosensitive ion channels<sup>67,76,87</sup> (but see<sup>88</sup>). Although the magnetite hypothesis is physically easy to explain<sup>76</sup>, other suggested effects of magnetic fields, such as ion-gating by moving ferritin complexes<sup>89,90</sup> and the MagR proposal of Qin et al.<sup>91</sup>, seem to be at odds with basic laws of physics<sup>69,72</sup>. To sum up, although magnetic particles have been found in many animals, there exists no independently confirmed ultrastructural evidence for the in situ presence of bacteria-like magnetite chains in sensory structures of any insect or vertebrate<sup>57,78,79</sup>.

### Radical-pair-based magnetoreception

The radical-pair hypothesis suggests that the quantum mechanics of electron spins (questions 2, 3) could form the basis of a magnetic compass sense<sup>71,92–95</sup>: a light-induced electron transfer reaction generates long-lived radical pairs, which can exist in singlet or triplet electronic spin-states. The coherent quantum mechanical interconversion between these two states is affected by the orientation of the sensor molecule relative to the geomagnetic field. This in turn affects the likelihood of forming a signalling state that could form the basis of a chemical magnetic compass sense that might enable birds to 'see' geomagnetic field parameters<sup>68,71,92,93,95,96</sup>. Here I summarize the key points of radical-pair-based magnetoreception; for fuller details I recommend a recently published review<sup>71</sup>.

At first sight, a radical-pair compass seems implausible: the energetic interaction of the geomagnetic field (25–65  $\mu\text{T}$ ) with a single molecule is more than a million times smaller than the molecule's thermal energy,  $k_B T$ , under physiological conditions<sup>71</sup>.  $k_B T$  is the energy associated with the ever-present random motions of molecules as they rotate, vibrate, and bump into one another<sup>71</sup>. Normally, a significant impact on the rate or yield of a chemical transformation is impossible unless the amount of energy supplied is at least comparable to  $k_B T$ . The tethering stone and fly analogy in Fig. 3 may help to explain why radical-pair reactions are different in this respect. Only when a system has previously been brought into an appropriate state far from equilibrium (the radical-pair state symbolized by the tethering stone), tiny interactions (the geomagnetic field symbolized by the fly) can have profound effects<sup>71</sup> (for details, see legend to Fig. 3; for formal arguments, see the recent review<sup>71</sup>).

The radical-pair mechanism is unquestionably genuine. There have been hundreds of laboratory studies of radical-pair reactions on which 1–100-mT magnetic fields have an effect<sup>71</sup>, and a model compound has been shown to be sensitive to Earth-strength magnetic fields<sup>94</sup>. However, it has not been demonstrated that this reaction scheme is responsible for animal magnetoreception<sup>71</sup> (questions 1–3). However, a substantial amount of correlative evidence supports this idea.

The magnetic compass of birds is an inclination compass, which detects the angle between the magnetic field lines and gravity rather than the polarity of the field<sup>1,58,97</sup>. The magnetic compass orientation of newts<sup>61,98</sup> and birds<sup>99</sup> depends on the wavelengths of light that are available during behavioural tests. This wavelength-dependence suggests that the eyes and/or pineal organ are involved in the magnetic compass. In birds, the pineal organ is not needed<sup>100</sup>, whereas pineal photoreceptor molecules seem to be essential for magnetic compass orientation in newts<sup>98</sup>.

Furthermore, radiofrequency magnetic fields disrupt magnetic compass orientation in several animals<sup>71,101–106</sup>. Radiofrequency fields can influence the spins of unpaired electrons in a radical pair and thus the probability of finding radical pairs in the singlet or triplet state<sup>71</sup>. To come back to the analogy shown in Fig. 3, it would be like exposing the granite block poised on its edge to a swarm of *Drosophila* hitting it from unpredictable and random directions before the bigger fly would get a chance to influence the fate of the block (Fig. 3c). By contrast, the radiofrequency fields are far too weak to break a chemical bond or physically move a magnetic particle.

A couple of cautionary notes: it has previously been predicted that time-dependent magnetic field effects should be specific to the Larmor frequency (the frequency with which electron spins precess in a plane

perpendicular to an external magnetic field if they are not influenced by any hyperfine interactions<sup>102</sup>. However, this prediction was based on several assumptions that are not true in any realistic biological molecule. A much broader band of frequencies should be disruptive to the magnetic compass<sup>71,107</sup>, and indeed they are<sup>103,106</sup>. Nevertheless, many studies have reported specific effects of exposing animals to Larmor-frequency fields. However, none of these studies provided measured broadband disturbance spectra. Therefore, substantial side bands and/or broadband background disturbances at other relevant frequencies might have occurred<sup>71,106</sup>. Even though radical-pair theory predicts sensitivity to radiofrequency fields, it is still not understood why the bird's magnetic compass is so extraordinarily sensitive to disruptive anthropogenic electromagnetic fields<sup>71,107</sup> (question 3).

What might be the identity of the light-dependent magnetic detectors (question 1)? Opsins cannot be radical-pair-based magnetoreceptors, because they use light energy to cause a conformational change and nowhere in their signalling cascade is a radical pair formed<sup>108</sup>. Cryptochrome proteins are the only photoreceptor molecules known in vertebrates that use light energy to form long-lived radical pairs<sup>71,93,109,110</sup>, and the radical-pair chemistry of cryptochromes have been shown to be magnetically sensitive<sup>71,110</sup>. Because the radical pair in cryptochromes forms between the protein and its flavin co-factor, only cryptochromes with their flavin co-factor present can be magnetically sensitive. Four different cryptochromes have been located in the retinas of migratory birds<sup>71,91,111–116</sup>, and whereas cryptochromes 1a, 1b, and 2 do not seem to bind flavin well<sup>117</sup>, cryptochrome 4 is a particularly attractive magnetosensory candidate because it binds flavin well<sup>91,116</sup>. Furthermore, cryptochrome 4 is located in double cones, which are two cones attached to each other that look at the same location in space and thus get very similar light input<sup>71,116</sup>. This should make it easier to separate magnetic field changes from light intensity and polarization changes<sup>71,116,118</sup>. Behavioural evidence from genetically modified *Drosophila* also supported the involvement of cryptochromes in magnetic sensing<sup>119,120</sup>, and theoretical studies of cryptochrome-like radical pairs have contributed much to our current understanding of how radical-pair-based magnetoreception could work<sup>68,70,71,95,96,107</sup>.

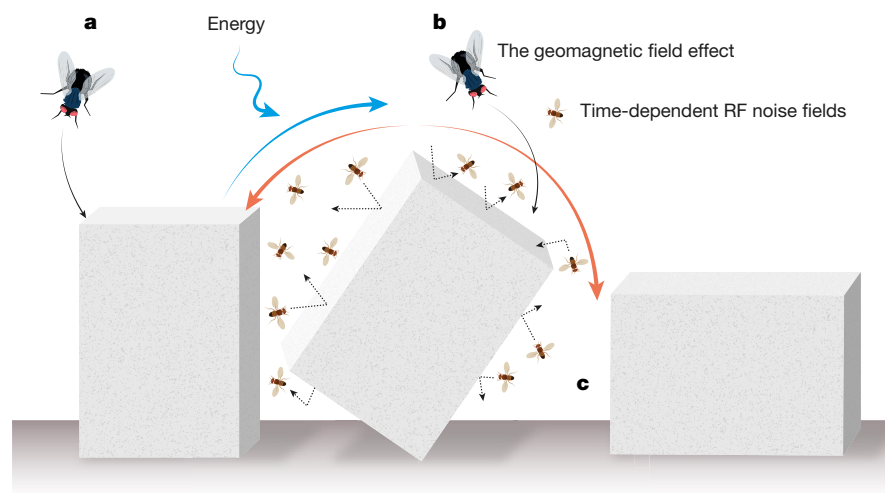
Can light-dependent magnetoreceptors work at night (questions 1, 4)? Theoretically the answer is yes; some light is always present. Even humans can see well enough to walk on an open field on a moonless overcast night because our rod photoreceptor cells are activated by only a few photons<sup>108</sup>. Light receptors responsible for light-dependent magnetoreception could also be activated by just a few photons. The key open questions are how the light-dependent magnetoreception mechanism collects sufficient reaction statistics to differentiate between magnetic directions under low light conditions, and how they separate changes in light intensity from magnetic field changes<sup>71,116,118</sup> (question 4).

Finally, brain activation patterns and a lesion study in night-migratory songbirds have shown that magnetic compass information is processed in Cluster N, a specific part of the thalamofugal visual brain pathway<sup>121–124</sup>. These findings strongly support the idea that light-dependent magnetoreception with primary detector molecules located in the eyes exists and that these birds perceive magnetic compass input as a visual cue<sup>71,121–124</sup>. An earlier claim<sup>125</sup> that the magnetic compass is located only in the bird's right eye has turned out to be incorrect<sup>126–128</sup>. Our knowledge about where in the brain magnetic information is processed in other animals is very sparse<sup>85,86</sup> (question 8).

In summary, there is much evidence that the magnetic compasses of night-migratory songbirds (and probably other animals) rely on the spin-chemistry of radical-pair reactions. This could be fundamentally important because, if radical-pair-based magnetoreception is real, it would firmly establish the emerging field of quantum biology and thereby reduce by 6–7 orders of magnitude the threshold for sensory detection of weak stimuli in biological systems<sup>68,71</sup>. To prove the existence of radical-pair-based magnetoreception, truly multidisciplinary collaborative approaches involving quantum physics, chemistry, computer simulation, and biochemistry in combination with molecular biology, neurobiology, and behavioural biology, will be needed (questions 1–4, 8).

### Can animals have more than one magnetic sense?

Traditionally, many have considered the magnetoreception hypotheses described above as mutually exclusive. This must not be. In fact, I would expect the magnetic map and magnetic compass senses to have



**Fig. 3 | A mechanical analogy of the radical-pair mechanism.** This analogy, originally designed by P. J. Hore, illustrates why a radical-pair reaction can be significantly affected by extremely small magnetic interactions. Imagine we have a heavy stone block at rest and ask whether a fly could tip it over (a). The answer, obviously, is no<sup>71</sup>. But suppose we have supplied the energy necessary to poise the stone on its sharp edge. Clearly, it would not be stable. It would very soon fall to the left or the right<sup>71</sup>. But what if a fly landed on its right-hand side while the block is teetering in this way (b)? Even though the energy imparted by the fly would be

minute, it could be enough to cause the block to fall to the right rather than the left<sup>71</sup>. Thus, tiny interactions can have profound effects, but only if a system has previously been brought into an appropriate state far from equilibrium<sup>71</sup>. In the context of radical-pair-based magnetoreception, the non-equilibrium state is the radical pair, the energy required to reach that state comes from a photon of light, and the fly is the static geomagnetic field<sup>71</sup>. Radiofrequency noise would be a bit like having a swarm of *Drosophila* (c) constantly bumping into the tethering stone block from all directions. Modified after ref. 71.

mutually distinct properties and mechanisms because a direction sensor should be insensitive to magnetic intensity and vice versa<sup>71</sup>.

Indeed, behavioural and brain activation data suggest that the magnetic compass of night-migratory songbirds is light-dependent and radical-pair-based<sup>93,99,101–103,106</sup> and is processed in Cluster N<sup>121–124</sup>. When Cluster N is lesioned, European robins can still use their sun compass and their star compass, but their magnetic compass no longer works<sup>124</sup>.

By contrast, when the ophthalmic branch of the trigeminal nerve (V1) is cut bilaterally, night-migratory songbirds seem unable to compensate for displacements; that is, their map sense is disrupted<sup>147,48</sup>, but their magnetic compass remains unaffected<sup>1,47,124</sup>. Furthermore, magnetic field-dependent neuronal activation has been documented in hindbrain regions innervated by V1<sup>129,130</sup>, and strong magnetic pulses, which should re-magnetize a magnetite-containing sensor, lead only to deflected headings in adult migratory birds that have established a map<sup>31,45,131</sup>. Both types of sensor also seem to exist in amphibians<sup>33,61</sup>, whereas there is evidence for a light-independent magnetoreceptor only in sea turtles and fish<sup>30,60</sup>. In conclusion, radical-pair-based and magnetic-particle-based magnetoreception mechanisms seem to exist side by side in several animals and may provide animals with magnetic compass and map information, respectively.

### Celestial cues and how they are sensed

Photoreceptor pigments in the eyes detect photons emitted from the sun and stars<sup>108</sup>, which can be used for orientation and navigation. Virtually every animal tested can derive compass information from the sun<sup>12,14,27,59,132,133</sup>. Night-migratory songbirds can also use the stars<sup>1,134</sup>.

### The sun compass and polarized light cues

The sun compass is learned and seems to rely only on the azimuthal direction of the sun<sup>132</sup>. To establish a sun compass that can be used for longer-distance orientation, young animals must observe and learn the path of the sun and must link the sun's azimuthal positions to their circadian clock<sup>14,15,132,135</sup>. Animals can adapt their compass responses as the sun's movements change with the season<sup>27,136</sup>.

The sun compass of many insects relies on detecting the polarized light pattern of the sky, which is generated when sunlight is scattered by molecules in the atmosphere<sup>14,15,135,137</sup>. Even though monarch butterflies can detect polarized light cues<sup>138</sup>, carefully controlled experiments found that, surprisingly, they seem to not use them for migratory orientation<sup>52</sup>. Whether vertebrates can detect polarized light remains unclear, with the best evidence for polarization vision coming from anchovies<sup>137,139</sup>.

The visual brain pathways are known in many animals, but where celestial orientation and navigation-relevant cues are specifically processed in the brains of vertebrates is much less clear<sup>32</sup>. In insects, sun compass information in the form of polarized light is detected in the dorsal rim area of the compound eye<sup>14,16,53,138</sup>. The information then passes through the medulla of the optic lobe<sup>14</sup> on its way to the central complex in the brain, where neurons coding for the e-vector axis of polarized light have been found<sup>14,15,53,135</sup>. Some central complex neurons in locusts even seem to represent matched filters to the natural polarization pattern, so that different cells respond to different orientations of the complete celestial polarization pattern across the dome of the sky<sup>15</sup>.

### The star compass

The star compass of night-migratory songbirds must be learned<sup>134</sup>. Night-migratory songbirds have no inherited knowledge of what the star patterns should look like. Instead, on the Northern Hemisphere, birds are born with the information to look for rotating light-dots in the sky and to interpret the centre of rotation as North<sup>134,140,141</sup>. More than seven clear nights seem to be needed in order for birds to establish their star compass<sup>134,140,141</sup>. Once this is established, birds learn the geometrical star patterns and thereafter no longer need to observe celestial rotation<sup>134,140,142</sup>. One fascinating open question is how animals detect the very slow rotation of the stars (question 14). Birds can learn the concept of a rotational centre<sup>143</sup>, but whether they actually see the slow rotation

or use a snapshot comparison mechanism remains unknown<sup>32,143</sup>. It is unclear whether nocturnal arthropods have a star compass, but they can at least use night-time celestial cues as beacons<sup>144</sup>.

### Olfactory cues and how they are sensed

Olfactory cues are volatile chemicals in air or soluble chemicals in water that are detected by receptor proteins<sup>145,146</sup>. The brain circuits responsible for olfaction in most vertebrates and many invertebrates are well understood<sup>145</sup>.

Odours play a very important role in homing: for example, of fish<sup>25,26,54,55</sup>, pigeons<sup>49,147</sup> and experienced pelagic seabirds<sup>10,49,55</sup>. Surprisingly, the ratios of several volatiles are highly stable within a 400 × 400-km<sup>2</sup> terrestrial area even across different seasons, and model pigeons could home using these ratios<sup>148</sup>. Odour-based maps are probably gradient maps that provide information only about the direction of displacement<sup>49</sup>. An inexperienced migrant cannot know how its destination thousands of kilometres away will smell. Thus, olfactory cues are likely to be most important during the homing and pinpointing-the-goal phases, but could also play a role during the long-distance phase in experienced navigators<sup>10,49</sup>. Insects can also use olfactory cues for navigation, but mainly over shorter distances, for instance when locating nests<sup>145</sup> or mating partners<sup>146</sup>.

### Landmarks

Landmarks can in principle be detected by any sense, and animals can use visual, olfactory, magnetic, and/or auditory landmarks<sup>1,10,13,14,26,34,49,56,62,149</sup>. Landmarks play an important role primarily during the last two phases of a navigational task (the homing and pinpointing-the-goal phases), but leading lines such as coastlines and mountain ranges can also be important as physical constraints during the long-distance phase<sup>1,10,13,14,26,34,39,49,56,149</sup>.

### Other cues

Some animals, such as charcoal beetles (*Melanophila* species), use infrared radiation (heat) detection to orient towards fires<sup>150</sup>. It has also been suggested that various animals can use very long-waved 'infrasound' to home<sup>151</sup>. It is, however, difficult to understand how animals with head sizes much smaller than the wavelength of infrasound could extract the needed directional information.

In addition to traditional navigational cues, some aerial and aquatic migrants should also consider the speed and direction of the currents in which they are moving<sup>17,19,20</sup>. Migratory insects are exquisitely adapted for choosing the most suitable days or nights and airstreams to optimize wind assistance in the preferred direction<sup>20,29</sup>. By doing so, they reach migration efficiencies that match those of migratory birds, even though their flight speeds are at least three times slower<sup>29</sup>. However, detection of the direction of flow by insects when they are embedded in it is not a trivial problem (question 15). Insects seem to detect micro-turbulence cues around their bodies and use these to detect flow direction<sup>152</sup>. Why can birds apparently not select favourable airflow layers as efficiently as insects<sup>20,29</sup>? I suspect that, in addition to their size<sup>152</sup>, their feather coating probably prevents micro-turbulence cues from reaching the somatosensory sensors in their skin, thereby preventing detection.

### Multisensory input

#### Evolutionary advantage of multisensory input

Traditionally, many studies aimed to show that one specific cue was used exclusively for navigation, and this focus has led to many apparent controversies and contradictions. Furthermore, many kinds of calibrations from one cue to another have been demonstrated<sup>1,153,154</sup>. In my opinion, there is no universally valid cue, single strategy, or fixed cue hierarchy that would enable 100% accurate navigation during all phases and in all situations. This view is strongly supported by a recent review, in which the authors attempted to route-fit single navigation mechanisms to tracking data from many free-flying migratory birds<sup>154</sup>. The authors concluded that no model exists that would fit all the data<sup>154</sup>. The relative cue importance seems to vary between species and phases

## Box 1

## Important open mechanistic questions

The questions listed below represent some of the most important open mechanistic questions for the next two decades of long-distance navigation and magnetoreception research.

1. How do the magnetic senses work on the biophysical, biochemical, and molecular levels?
2. Does quantum biology exist (that is, is magnetic sensing truly quantum in at least some animals)?
3. What are the explanation for and ecological consequences of the extraordinary sensitivity of the bird's magnetic compass to disruptive anthropogenic electromagnetic fields?
4. How does the light-dependent magnetoreception mechanism distinguish between changes in light intensity and magnetic direction, and how does it collect enough reaction statistics to detect magnetic directions under low light conditions?
5. Do some animals use electromagnetic induction to detect the geomagnetic field?
6. Do magnetic particles exist inside cells at consistent and specific locations in many individuals of any migratory animal, and are the particles associated with the nervous system?
7. How, if at all, can slow-moving animals distinguish the spatial magnetic signal from temporal geomagnetic field variation to allow for a magnetic map with a resolution below 10–30 km?
8. Where and how is magnetic information sensed and processed?
9. Where in the brain, and how, is multisensory navigational information integrated and weighted?
10. How do processing strategies in the nervous system transition between the different phases?
11. How does the brain deal with conflicting and/or incomplete information, and does this depend on the ecological conditions and/or the navigational phase?
12. Do place and grid cell equivalents exist as neural correlates of the map over scales of kilometres or even thousands of kilometres, and, if yes, which cues contribute to their establishment?
13. Do equivalents of head direction cells exist that code for celestial and/or magnetic compass direction on a regional or global scale?
14. How is the very slow rotation of the stars detected?
15. How do small animals moving in air or water detect the direction of flow even though they are embedded in the flowing medium themselves?
16. Which genes trigger migration behaviour and/or code for migratory direction and distance?
17. Exactly what cues signal to an animal that it should start migrating or that it has reached its destination and should terminate migration?
18. What determines when an animal switches from one navigational phase to the next?
19. How is longitude (east–west) position determined on a regional or even global scale?
20. How does the pinpointing-the-goal phase work in a monarch butterfly or Bogong moth, which can pinpoint their very specific wintering locations even though they have never been there before?

and with ecological context<sup>1,59,153–155</sup> (question 11). This is not very surprising, as animals that can use several navigation strategies and integrate information from all potentially relevant cues will be more versatile and therefore have a long-term evolutionary advantage over animals that use only a single strategy and cue. Understanding multisensory integration in the animals' brain will thus be key to understanding animal navigation.

## Multisensory cue integration in the brain

In birds, the hippocampus and the caudolateral nidopallium (NCL), which receive input from all sensory modalities, could be involved in multisensory integration, in the weighting of navigational cues, and/or in deciding to fly in a particular direction at any given moment in time<sup>32</sup>. In insects, the integration of multisensory navigational cues and decision-making is most likely to happen in the central complex<sup>16</sup>.

Once the integrative centres in the brain have been identified (question 9), we can investigate how animals estimate the reliability of each navigational cue, how animals use these estimates, and whether animals take an estimated-reliability-weighted average or use a winner-takes-all strategy (question 9). Maybe cues that are estimated to be less reliable than a certain threshold will be ignored completely. Consequently, unclear or unnatural stimuli provided during a scientific experiment might be ignored, even if an animal could in principle sense them. For instance<sup>107</sup>, if anthropogenic radiofrequency fields, a source of noise not present until about 100 years ago, add noise to the perception of magnetic fields, magnetic cues could be ignored even though the noise is not strong enough to entirely mask the static geomagnetic signal (question 13).

## Neural representations of map and compass

The rodent hippocampus contains place cells, which define a specific location within a small arena, and head direction cells, which represent the animal's current heading<sup>50</sup>. Furthermore, the entorhinal cortex contains grid cells, which fire at node-points in a repetitive triangular array covering the entire available surface<sup>156</sup>. Grid cells might define distances<sup>50,156</sup>. These fascinating cell types are highly likely to be neural representations of location and direction during the pinpointing-the-goal phase, as these cell types are established relative to prominent local landmarks<sup>50,156</sup>. In contrast to the extensive knowledge about short-distance navigation in rats, mice, and fruit bats, very little is known about long-distance navigation mechanisms in mammals. Do similar cell types exist that define direction (compass information) and location (map information) during the homing and long-distance phases of a navigational task (question 12)? If so, their responses would need to be established relative to global cues such as celestial bodies or the geomagnetic field (question 13), because long-distance migrants and homing animals can determine direction and location in unfamiliar places. Furthermore, during the pinpointing-the-goal phase, the spatial coding cells of many animals will need to define three-dimensional space. Recently, place, grid, head, and goal direction cells defined in three-dimensional space were found in flying Egyptian fruit bats<sup>50,51</sup>. Compass neurons also exist in the central complex of migratory insects (see above). Map concepts—let alone map neurons—are very controversial among insect researchers<sup>13,14,157</sup>.

## Key open questions for the next two decades

Despite substantial advances in our understanding of long-distance animal navigation and magnetoreception over the last two decades, many fascinating questions remain unanswered. The twenty questions in Box 1 are a summary of the most important mechanistic questions that arose from preparing this review (their order does not indicate relative importance). To answer many of them, a long-term collaborative effort combining new multidisciplinary approaches from quantum mechanics and biophysics, via molecular biology, biochemistry, neurobiology, and genetics all the way to perception and behaviour of the intact animal will be required. These will be exciting times in the field.

Received: 9 July 2017; Accepted: 13 April 2018;

Published online 6 June 2018.

1. Wiltschko, R. & Wiltschko, W. *Magnetic Orientation in Animals* (Springer, Berlin, 1995).

**This book is an exhaustive account of almost all studies related to magnetoreception in any animal published before 1995, and it is a valuable historical account of the early achievements in the field.**

2. Berthold, P. A comprehensive theory for the evolution, control and adaptability of avian migration. *Ostrich* **70**, 1–11 (1999).

3. Mouritsen, H. in *Avian Migration* (eds Berthold, P., Gwinner, E. & Sonnenschein, E.) 493–513 (Springer, Berlin, 2003).
4. Mouritsen, H. in *Sturkie's Avian Physiology* (ed. Scanes, C.) 113–133 (Elsevier, Amsterdam, 2015).
5. Schmaljohann, H., Fox, J. W. & Bairlein, F. Phenotypic response to environmental cues, orientation and migration costs in songbirds flying halfway around the world. *Anim. Behav.* **84**, 623–640 (2012).
6. Salewski, V., Bairlein, F. & Leisler, B. Recurrence of some palaearctic migrant passerine species in West Africa. *Ring. Migr.* **20**, 29–30 (2000).
7. Gill, R. E. Jr et al. Extreme endurance flights by landbirds crossing the Pacific Ocean: ecological corridor rather than barrier? *Proc. R. Soc. Lond. B* **276**, 447–457 (2009).
8. Egevang, C. et al. Tracking of Arctic terns *Sterna paradisaea* reveals longest animal migration. *Proc. Natl Acad. Sci. USA* **107**, 2078–2081 (2010).
9. Jouventin, P. & Weimerskirch, H. Satellite tracking of wandering albatrosses. *Nature* **343**, 746–748 (1990).
10. Gagliardo, A. et al. Oceanic navigation in Cory's shearwaters: evidence for a crucial role of olfactory cues for homing after displacement. *J. Exp. Biol.* **216**, 2798–2805 (2013).  
**This paper convincingly showed that olfactory information is essential for long-distance homing in Cory's shearwaters, because birds fitted with satellite transmitters and released about 800 km from home with their olfactory nerves cut wandered aimlessly around the Atlantic Ocean, whereas shearwaters with intact olfactory nerves but with cut ophthalmic branches of the trigeminal nerves went straight home.**
11. Brower, L. Monarch butterfly orientation: missing pieces of a magnificent puzzle. *J. Exp. Biol.* **199**, 93–103 (1996).
12. Mouritsen, H. & Frost, B. J. Virtual migration in tethered flying monarch butterflies reveals their orientation mechanisms. *Proc. Natl Acad. Sci. USA* **99**, 10162–10166 (2002).
13. Zeil, J. Visual homing: an insect perspective. *Curr. Opin. Neurobiol.* **22**, 285–293 (2012).
14. Wehner, R., Cheng, K. & Cruse, H. in *The New Visual Neurosciences* 1153–1164 (MIT Press, Cambridge, 2014).
15. Bech, M., Homberg, U. & Pfeiffer, K. Receptive fields of locust brain neurons are matched to polarization patterns of the sky. *Curr. Biol.* **24**, 2124–2129 (2014).  
**This elegant electrophysiological paper used the scientific advantage of the simplicity of the insect brain to show that some neurons in the central complex of locusts seem to be matched filters to the natural polarization pattern, so that different cells respond to different orientations of the complete celestial polarization pattern across the dome of the sky, and that these neurons can differentiate between solar and antisolar directions based only on the polarization pattern.**
16. Heinze, S. Neuroethology: unweaving the senses of direction. *Curr. Biol.* **25**, R1034–R1037 (2015).
17. Chapman, J. W., Reynolds, D. R. & Wilson, K. Long-range seasonal migration in insects: mechanisms, evolutionary drivers and ecological consequences. *Ecol. Lett.* **18**, 287–302 (2015).
18. Warrant, E. et al. The Australian bogong moth *Agrotis infusa*: a long-distance nocturnal navigator. *Front. Behav. Neurosci.* **10**, 77 (2016).
19. Chapman, J. W. et al. Wind selection and drift compensation optimize migratory pathways in a high-flying moth. *Curr. Biol.* **18**, 514–518 (2008).  
**This paper was the first to show that high-flying insects are not at the mercy of the wind but that they actively orient themselves in mid-air and that they choose favourable airstreams that enable them to perform directed migration in spring and return migration in autumn; this paper therefore also disproved the 'pied piper' hypothesis that high-flying insects were blown in random directions.**
20. Chapman, J. W. et al. Flight orientation behaviors promote optimal migration trajectories in high-flying insects. *Science* **327**, 682–685 (2010).
21. Hu, G. et al. Mass seasonal bioflows of high-flying insect migrants. *Science* **354**, 1584–1587 (2016).
22. Lohmann, K. J., Cain, S. D., Dodge, S. A. & Lohmann, C. M. F. Regional magnetic fields as navigational markers for sea turtles. *Science* **294**, 364–366 (2001).
23. Putman, N. F. et al. Evidence for geomagnetic imprinting as a homing mechanism in Pacific salmon. *Curr. Biol.* **23**, 312–316 (2013).  
**This elegant paper used fisheries data and information on geomagnetic field drift to demonstrate that Pacific salmon returning to spawn had imprinted on the geomagnetic parameters of their natal river mouth before they left the area years earlier.**
24. Brothers, J. R. & Lohmann, K. J. Evidence for geomagnetic imprinting and magnetic navigation in the natal homing of sea turtles. *Curr. Biol.* **25**, 392–396 (2015).
25. Bett, N. N. & Hinch, S. G. Olfactory navigation during spawning migrations: a review and introduction of the hierarchical navigation hypothesis. *Biol. Rev. Camb. Philos. Soc.* **91**, 728–759 (2016).
26. Gerlach, G., Atema, J., Kingsford, M. J., Black, K. P. & Miller-Sims, V. Smelling home can prevent dispersal of reef fish larvae. *Proc. Natl Acad. Sci. USA* **104**, 858–863 (2007).  
**This paper used genetic fingerprinting and behavioural tests to elegantly demonstrate that returning reef fish larvae are attracted to the odour of their natal reef, that they can discriminate this odour from the odour of other reefs, and that this olfactory imprinting on their natal reef might help explain the high levels of retention and speciation in coral reefs.**
27. Mouritsen, H., Atema, J., Kingsford, M. J. & Gerlach, G. Sun compass orientation helps coral reef fish larvae return to their natal reef. *PLoS One* **8**, e66039 (2013).
28. Bottesch, M. et al. A magnetic compass that might help coral reef fish larvae return to their natal reef. *Curr. Biol.* **26**, R1266–R1267 (2016).
29. Alerstam, T. et al. Convergent patterns of long-distance nocturnal migration in noctuid moths and passerine birds. *Proc. R. Soc. Lond. B* **278**, 3074–3080 (2011).
30. Lohmann, K. J., Lohmann, C. M. F., Brothers, J. R. & Putman, N. F. in *The Biology of Sea Turtles* (eds Wyneken, J., Lohmann, K. J. & Musick, J. A.) vol. **3**, 59–77 (CRC Press, Boca Raton, 2013).
31. Holland, R. A. True navigation in birds: from quantum physics to global migration. *J. Zool. (Lond.)* **293**, 1–15 (2014).
32. Mouritsen, H., Heyers, D. & Güntürkün, O. The neural basis of long-distance navigation in birds. *Annu. Rev. Physiol.* **78**, 133–154 (2016).
33. Phillips, J. B. Two magnetoreception pathways in a migratory salamander. *Science* **233**, 765–767 (1986).
34. Guilford, T. & Biro, D. Route following and the pigeon's familiar area map. *J. Exp. Biol.* **217**, 169–179 (2014).
35. Griffin, D. R. Bird navigation. *Biol. Rev. Camb. Philos. Soc.* **27**, 359–400 (1952).
36. Perdeck, A. C. Two types of orientation in migrating *Sturnus vulgaris* and *Fringilla coelebs* as revealed by displacement experiments. *Ardea* **46**, 1–37 (1958).
37. Chernetsov, N., Kishkinev, D. & Mouritsen, H. A long-distance avian migrant compensates for longitudinal displacement during spring migration. *Curr. Biol.* **18**, 188–190 (2008).
38. Chernetsov, N. et al. Migratory Eurasian reed warblers can use magnetic declination to solve the longitude problem. *Curr. Biol.* **27**, 2647–2651.e2 (2017).  
**This paper showed that adult, but not juvenile, Eurasian reed warblers can use magnetic declination—which requires two compasses—to correct for a virtual magnetic displacement from Kaliningrad to Scotland and therefore suggest that many bird species in Europe and North America could use magnetic declination to solve the enigmatic longitude problem.**
39. Mouritsen, H. et al. An experimental displacement and over 50 years of tag-recoveries show that monarch butterflies are not true navigators. *Proc. Natl Acad. Sci. USA* **110**, 7348–7353 (2013).
40. Lugo Ramos, J. S., Delmore, K. E. & Liedvogel, M. Candidate genes for migration do not distinguish migratory and non-migratory birds. *J. Comp. Physiol.* **203**, 383–397 (2017).
41. Lohmann, K. J., Lohmann, C. M. F. & Putman, N. F. Magnetic maps in animals: nature's GPS. *J. Exp. Biol.* **210**, 3697–3705 (2007).
42. Mouritsen, H. & Mouritsen, O. A mathematical expectation model for bird navigation based on the clock-and-compass strategy. *J. Theor. Biol.* **207**, 283–291 (2000).
43. Thorup, K. et al. Evidence for a navigational map stretching across the continental U.S. in a migratory songbird. *Proc. Natl Acad. Sci. USA* **104**, 18115–18119 (2007).
44. Deutschlander, M. E., Phillips, J. B. & Munro, U. Age-dependent orientation to magnetically-simulated geographic displacements in migratory Australian silvereyes (*Zosterops l. lateralis*). *Wilson J. Ornithol.* **124**, 467–477 (2012).
45. Holland, R. A. & Helm, B. A strong magnetic pulse affects the precision of departure direction of naturally migrating adult but not juvenile birds. *J. R. Soc. Interface* **10**, 20121047 (2013).
46. Phillips, J. B., Freake, M. J., Fischer, J. H. & Borland, C. Behavioral titration of a magnetic map coordinate. *J. Comp. Physiol.* **188**, 157–160 (2002).
47. Kishkinev, D., Chernetsov, N., Heyers, D. & Mouritsen, H. Migratory reed warblers need intact trigeminal nerves to correct for a 1,000 km eastward displacement. *PLoS One* **8**, e65847 (2013).
48. Kishkinev, D., Chernetsov, N., Pakhomov, A., Heyers, D. & Mouritsen, H. Eurasian reed warblers compensate for virtual magnetic displacement. *Curr. Biol.* **25**, R822–R824 (2015).
49. Gagliardo, A. Forty years of olfactory navigation in birds. *J. Exp. Biol.* **216**, 2165–2171 (2013).
50. Geva-Sagiv, M., Las, L., Yovel, Y. & Ulanovsky, N. Spatial cognition in bats and rats: from sensory acquisition to multiscale maps and navigation. *Nat. Rev. Neurosci.* **16**, 94–108 (2015).
51. Sarel, A., Finkelstein, A., Las, L. & Ulanovsky, N. Vectorial representation of spatial goals in the hippocampus of bats. *Science* **355**, 176–180 (2017).  
**This paper discovered a new type of spatial cell in the hippocampus of free-flying Egyptian fruit bats that is essential for navigational tasks—namely cells coding for the direction to a goal relative to an animal's current heading.**
52. Stalleicken, J. et al. Do monarch butterflies use polarized skylight for migratory orientation? *J. Exp. Biol.* **208**, 2399–2408 (2005).
53. Heinze, S. & Reppert, S. M. Sun compass integration of skylight cues in migratory monarch butterflies. *Neuron* **69**, 345–358 (2011).
54. Scholz, A. T., Horral, R. M., Cooper, J. C. & Hasler, A. D. Imprinting to chemical cues: the basis for home stream selection in salmon. *Science* **192**, 1247–1249 (1976).
55. DeBose, J. L. & Nevitt, G. A. The use of odors at different spatial scales: comparing birds with fish. *J. Chem. Ecol.* **34**, 867–881 (2008).
56. Radford, C. A., Stanley, J. A., Simpson, S. D. & Jeffs, A. G. Juvenile coral reef fish use sound to locate habitats. *Coral Reefs* **30**, 295–305 (2011).
57. Mouritsen, H. in *Neurosciences—From Molecule to Behavior: A University Textbook* (eds Galizia, C. G. & Lledo, P.-M.) 427–443 (Springer, Heidelberg, 2013).
58. Wiltschko, W. & Wiltschko, R. Magnetic compass of European robins. *Science* **176**, 62–64 (1972).

59. Cochran, W. W., Mouritsen, H. & Wikelski, M. Migrating songbirds recalibrate their magnetic compass daily from twilight cues. *Science* **304**, 405–408 (2004).
60. Lohmann, K. J. & Lohmann, C. A light-independent magnetic compass in the leatherback sea turtle. *Biol. Bull.* **185**, 149–151 (1993).
61. Phillips, J. B. & Borland, S. C. Behavioral evidence for use of a light-dependent magnetoreception mechanism by a vertebrate. *Nature* **359**, 142–144 (1992).
62. Dennis, T. E., Rayner, M. J. & Walker, M. M. Evidence that pigeons orient to geomagnetic intensity during homing. *Proc. R. Soc. Lond. B* **274**, 1153–1158 (2007).
63. Guerra, P. A., Gegear, R. J. & Reppert, S. M. A magnetic compass aids monarch butterfly migration. *Nat. Commun.* **5**, 4164 (2014).
64. Komolkin, A. V. et al. Theoretically possible spatial accuracy of geomagnetic maps used by migrating animals. *J. R. Soc. Interface* **14**, 20161002 (2017).
65. Bazylinski, D. A. & Frankel, R. B. Magnetosome formation in prokaryotes. *Nat. Rev. Microbiol.* **2**, 217–230 (2004).
66. Begall, S., Malkemper, E. P., Červený, J., Němec, P. & Burda, H. Magnetic alignment in mammals and other animals. *Mamm. Biol.* **78**, 10–20 (2013).
67. Kirschvink, J. L., Winklhofer, M. & Walker, M. M. Biophysics of magnetic orientation: strengthening the interface between theory and experimental design. *J. R. Soc. Interface* **7**, S179–S191 (2010).
68. Solov'yov, I., Hore, P. J., Ritz, T. & Schulten, K. in *Quantum Effects in Biology* 218–236 (Cambridge Univ. Press, Cambridge, 2014).
69. Meister, M. Physical limits to magnetogenetics. *eLife* **5**, e17210 (2016).
70. Kattinig, D. R., Sowa, J. K., Solov'yov, I. A. & Hore, P. J. Electron spin relaxation can enhance the performance of a cryptochrome-based magnetic compass sensor. *New J. Phys.* **18**, 063007 (2016).
71. Hore, P. J. & Mouritsen, H. The radical-pair mechanism of magnetoreception. *Annu. Rev. Biophys.* **45**, 299–344 (2016).
- This tutorial review summarizes in detail all aspects of the radical-pair mechanism and the evidence for and against it as a magnetoreception mechanism, and aims to provide a must-read text for new scientists entering this field by explaining the biological aspects of the mechanism to physicists and chemists and the physicochemical and quantum mechanical aspects to biologists.**
72. Winklhofer, M. & Mouritsen, H. A magnetic protein compass? Preprint at <https://www.biorxiv.org/content/early/2016/12/15/094607> (2016).
73. Paulin, M. G. Electroreception and the compass sense of sharks. *J. Theor. Biol.* **174**, 325–339 (1995).
74. Rosenblum, B., Jungerman, R. L. & Longfellow, L. in *Magnetite Biomineralization and Magnetoreception in Organisms* 223–232 (Plenum, New York, 1985).
75. Uebe, R. & Schüler, D. Magnetosome biogenesis in magnetotactic bacteria. *Nat. Rev. Microbiol.* **14**, 621–637 (2016).
76. Winklhofer, M. & Kirschvink, J. L. A quantitative assessment of torque-transducer models for magnetoreception. *J. R. Soc. Interface* **7**, S273–S289 (2010).
77. Shaw, J. et al. Magnetic particle-mediated magnetoreception. *J. R. Soc. Interface* **12**, 20150499 (2015).
78. Treiber, C. D. et al. Clusters of iron-rich cells in the upper beak of pigeons are macrophages not magnetosensitive neurons. *Nature* **484**, 367–370 (2012).
79. Mouritsen, H. Sensory biology: Search for the compass needles. *Nature* **484**, 320–321 (2012).
80. Eder, S. H. et al. Magnetic characterization of isolated candidate vertebrate magnetoreceptor cells. *Proc. Natl Acad. Sci. USA* **109**, 12022–12027 (2012).
81. Edelman, N. B. et al. No evidence for intracellular magnetite in putative vertebrate magnetoreceptors identified by magnetic screening. *Proc. Natl Acad. Sci. USA* **112**, 262–267 (2015).
- This paper, together with Ref. 78, demonstrated that structures previously suggested to be strong candidates as magnetic-particle-based magnetoreceptors were dirt or non-magnetic iron accumulations, emphasizing that, to be considered as serious magnetoreception sensor candidates, magnetic particles must be proven to be located inside cells in exactly the same location and associated with nerve tissue in many individuals of the same species.**
82. Walker, M. M. et al. Structure and function of the vertebrate magnetic sense. *Nature* **390**, 371–376 (1997).
83. Fleissner, G. et al. Ultrastructural analysis of a putative magnetoreceptor in the beak of homing pigeons. *J. Comp. Neurol.* **458**, 350–360 (2003).
84. Wu, L.-Q. & Dickman, J. D. Neural correlates of a magnetic sense. *Science* **336**, 1054–1057 (2012).
85. Němec, P., Altmann, J., Marhold, S., Burda, H. & Oelschläger, H. H. A. Neuroanatomy of magnetoreception: the superior colliculus involved in magnetic orientation in a mammal. *Science* **294**, 366–368 (2001).
86. Burger, T. et al. Changing and shielded magnetic fields suppress c-Fos expression in the navigation circuit: input from the magnetosensory system contributes to the internal representation of space in a subterranean rodent. *J. R. Soc. Interface* **7**, 1275–1292 (2010).
87. Johnsen, S. & Lohmann, K. J. The physics and neurobiology of magnetoreception. *Nat. Rev. Neurosci.* **6**, 703–712 (2005).
88. Cadiou, H. & McNaughton, P. A. Avian magnetite-based magnetoreception: a physiologist's perspective. *J. R. Soc. Interface* **7**, S193–S205 (2010).
89. Stanley, S. A., Sauer, J., Kane, R. S., Dordick, J. S. & Friedman, J. M. Remote regulation of glucose homeostasis in mice using genetically encoded nanoparticles. *Nat. Med.* **21**, 92–98 (2015).
90. Wheeler, M. A. et al. Genetically targeted magnetic control of the nervous system. *Nat. Neurosci.* **19**, 756–761 (2016).
91. Qin, S. et al. A magnetic protein biocompass. *Nat. Mater.* **15**, 217–226 (2016).
92. Schulten, K., Swenberg, C. E. & Weller, A. A biomagnetic sensory mechanism based on magnetic field modulated coherent electron spin motion. *Z. Phys. Chem.* **111**, 1–5 (1978).
- This hardcore theoretical physics paper formulated the radical-pair hypothesis of magnetoreception for the first time, and it is now clear that it was decades ahead of its time.**
93. Ritz, T., Adem, S. & Schulten, K. A model for photoreceptor-based magnetoreception in birds. *Biophys. J.* **78**, 707–718 (2000).
94. Maeda, K. et al. Chemical compass model of avian magnetoreception. *Nature* **453**, 387–390 (2008).
- This paper proved that a radical-pair mechanism is fundamentally able to detect Earth-strength magnetic fields, as the authors synthesized a model compound in which they could directly observe that the photochemistry of a radical-pair mechanism was sensitive to Earth-strength magnetic fields.**
95. Hiscock, H. G. et al. The quantum needle of the avian magnetic compass. *Proc. Natl Acad. Sci. USA* **113**, 4634–4639 (2016).
96. Solov'yov, I. A., Mouritsen, H. & Schulten, K. Acuity of a cryptochrome and vision-based magnetoreception system in birds. *Biophys. J.* **99**, 40–49 (2010).
97. Schwarze, S. et al. Migratory blackcaps can use their magnetic compass at 5 degrees inclination, but are completely random at 0 degrees inclination. *Sci. Rep.* **6**, 33805 (2016).
98. Phillips, J. B., Deutschlander, M. E., Freake, M. J. & Borland, S. C. The role of extraocular photoreceptors in newt magnetic compass orientation: parallels between light-dependent magnetoreception and polarized light detection in vertebrates. *J. Exp. Biol.* **204**, 2543–2552 (2001).
99. Wiltschko, W., Munro, U., Ford, H. & Wiltschko, R. Red light disrupts magnetic orientation of migratory birds. *Nature* **364**, 525–527 (1993).
100. Schneider, T., Thalau, H. P., Semm, P. & Wiltschko, W. Melatonin is crucial for the migratory orientation of pied flycatchers *Ficedula hypoleuca pallas*. *J. Exp. Biol.* **194**, 255–262 (1994).
101. Ritz, T., Thalau, P., Phillips, J. B., Wiltschko, R. & Wiltschko, W. Resonance effects indicate a radical-pair mechanism for avian magnetic compass. *Nature* **429**, 177–180 (2004).
102. Ritz, T. et al. Magnetic compass of birds is based on a molecule with optimal directional sensitivity. *Biophys. J.* **96**, 3451–3457 (2009).
103. Engels, S. et al. Anthropogenic electromagnetic noise disrupts magnetic compass orientation in a migratory bird. *Nature* **509**, 353–356 (2014).
- This paper demonstrated in a massive series of reproducible, double-blinded experiments that anthropogenic electromagnetic fields in the low megahertz range and with an intensity 1,000 times lower than the current WHO guideline levels disrupt the magnetic compass sense of a night-migratory songbird; this strongly suggests that a quantum mechanical mechanism is responsible for magnetic compass sensing in these birds.**
104. Kavokin, K. et al. Magnetic orientation of garden warblers (*Sylvia borin*) under 1.4 MHz radiofrequency magnetic field. *J. R. Soc. Interface* **11**, 20140451 (2014).
105. Malkemper, E. P. et al. Magnetoreception in the wood mouse (*Apodemus sylvaticus*): influence of weak frequency-modulated radio frequency fields. *Sci. Rep.* **5**, 9917 (2015).
106. Schwarze, S. et al. Weak broadband electromagnetic fields are more disruptive to magnetic compass orientation in a night-migratory songbird (*Erithacus rubecula*) than strong narrow-band fields. *Front. Behav. Neurosci.* **10**, 55 (2016).
107. Hiscock, H. G., Mouritsen, H., Manolopoulos, D. E. & Hore, P. J. Disruption of magnetic compass orientation in migratory birds by radiofrequency electromagnetic fields. *Biophys. J.* **113**, 1475–1484 (2017).
108. Björn, L. O. *Photobiology: The Science of Light and Life* (Springer, New York, 2015).
109. Liedvogel, M. et al. Chemical magnetoreception: bird cryptochrome 1a is excited by blue light and forms long-lived radical-pairs. *PLoS One* **2**, e1106 (2007).
110. Maeda, K. et al. Magnetically sensitive light-induced reactions in cryptochrome are consistent with its proposed role as a magnetoreceptor. *Proc. Natl Acad. Sci. USA* **109**, 4774–4779 (2012).
111. Mouritsen, H. et al. Cryptochromes and neuronal-activity markers colocalize in the retina of migratory birds during magnetic orientation. *Proc. Natl Acad. Sci. USA* **101**, 14294–14299 (2004).
112. Liedvogel, M. & Mouritsen, H. Cryptochromes—a potential magnetoreceptor: what do we know and what do we want to know? *J. R. Soc. Interface* **7**, S147–S162 (2010).
113. Niessner, C. et al. Avian ultraviolet/violet cones identified as probable magnetoreceptors. *PLoS One* **6**, e20091 (2011).
114. Nießner, C. et al. Seasonally changing cryptochrome 1b expression in the retinal ganglion cells of a migrating passerine bird. *PLoS One* **11**, e0150377 (2016).
115. Bolte, P. et al. Localisation of the putative magnetoreceptive protein cryptochrome 1b in the retinae of migratory birds and homing pigeons. *PLoS One* **11**, e0147819 (2016).
116. Günther, A. et al. Double-cone localization and seasonal expression pattern suggest a role in magnetoreception for European robin cryptochrome 4. *Curr. Biol.* **28**, 211–223.e4 (2018).
- This paper suggests that cryptochrome 4 of night-migratory songbirds is a particularly strong candidate as the light-dependent magnetoreceptive protein because Cry4, in the retina, is exclusively expressed in the outer segments of the double cone and long-wavelength single cone**

- photoreceptor cells, and is more strongly expressed in the migratory season in migratory birds, whereas no seasonal differences are observed in non-migratory birds.**
117. Kutta, R. J., Archipowa, N., Johannissen, L. O., Jones, A. R. & Scrutton, N. S. Vertebrate cryptochromes are vestigial flavoproteins. *Sci. Rep.* **7**, 44906 (2017).
  118. Worster, S., Mouritsen, H. & Hore, P. J. A light-dependent magnetoreception mechanism insensitive to light intensity and polarization. *J. R. Soc. Interface* **14**, 20170405 (2017).
  119. Gegear, R. J., Casselman, A., Waddell, S. & Reppert, S. M. Cryptochrome mediates light-dependent magnetosensitivity in *Drosophila*. *Nature* **454**, 1014–1018 (2008).
  120. Fedele, G., Green, E. W., Rosato, E. & Kyriacou, C. P. An electromagnetic field disrupts negative geotaxis in *Drosophila* via a CRY-dependent pathway. *Nat. Commun.* **5**, 4391 (2014).
  121. Mouritsen, H., Feenders, G., Liedvogel, M., Wada, K. & Jarvis, E. D. Night-vision brain area in migratory songbirds. *Proc. Natl Acad. Sci. USA* **102**, 8339–8344 (2005).
  122. Heyers, D., Manns, M., Luksch, H., Güntürkün, O. & Mouritsen, H. A visual pathway links brain structures active during magnetic compass orientation in migratory birds. *PLoS One* **2**, e937 (2007).
  123. Zapka, M., Heyers, D., Liedvogel, M., Jarvis, E. D. & Mouritsen, H. Night-time neuronal activation of Cluster N in a day- and night-migrating songbird. *Eur. J. Neurosci.* **32**, 619–624 (2010).
  124. Zapka, M. et al. Visual but not trigeminal mediation of magnetic compass information in a migratory bird. *Nature* **461**, 1274–1277 (2009).  
**This paper demonstrates that Cluster N processes light-dependent magnetic compass information in night-migratory songbirds, because Cluster N-lesioned birds could still use their sun and star compasses but not their magnetic compass, and because Cluster N is part of the thalamofugal visual pathway in night-migratory songbirds<sup>129</sup>.**
  125. Wiltshcko, W., Traudt, J., Güntürkün, O., Prior, H. & Wiltshcko, R. Lateralization of magnetic compass orientation in a migratory bird. *Nature* **419**, 467–470 (2002).
  126. Hein, C. M., Engels, S., Kishkinev, D. & Mouritsen, H. Robins have a magnetic compass in both eyes. *Nature* **471**, E11–E12 (2011).
  127. Wiltshcko, W., Traudt, J., Güntürkün, O., Prior, H. & Wiltshcko, R. Wiltshcko et al. reply. *Nature* **471**, E12–E13 (2011).
  128. Engels, S., Hein, C. M., Lefeldt, N., Prior, H. & Mouritsen, H. Night-migratory songbirds possess a magnetic compass in both eyes. *PLOS One* **7**, e43271 (2012).
  129. Heyers, D., Zapka, M., Hoffmeister, M., Wild, J. M. & Mouritsen, H. Magnetic field changes activate the trigeminal brainstem complex in a migratory bird. *Proc. Natl Acad. Sci. USA* **107**, 9394–9399 (2010).
  130. Elbers, D., Bulte, M., Bairlein, F., Mouritsen, H. & Heyers, D. Magnetic activation in the brain of the migratory northern wheatear (*Oenanthe oenanthe*). *J. Comp. Physiol.* **203**, 591–600 (2017).
  131. Munro, U., Munro, J. A., Phillips, J. B., Wiltshcko, R. & Wiltshcko, W. Evidence for a magnetite-based navigational ‘map’ in birds. *Naturwissenschaften* **84**, 26–28 (1997).
  132. Wiltshcko, W., Wiltshcko, R. & Keeton, W. T. Effects of a ‘permanent’ clock-shift on the orientation of young homing pigeons. *Behav. Ecol. Sociobiol.* **1**, 229–243 (1976).
  133. Schmidt-Koenig, K., Ganzhorn, J. U. & Ranvaud, R. in *Orientation in Birds* 1–15 (Birkhäuser, Basel, 1991).
  134. Emlen, S. T. The stellar-orientation system of a migratory bird. *Sci. Am.* **233**, 102–111 (1975).
  135. Heinze, S. & Homberg, U. Maplike representation of celestial E-vector orientations in the brain of an insect. *Science* **315**, 995–997 (2007).
  136. Wiltshcko, R., Walker, M. & Wiltshcko, W. Sun-compass orientation in homing pigeons: compensation for different rates of change in azimuth? *J. Exp. Biol.* **203**, 889–894 (2000).
  137. Horváth, G. (Ed.) *Polarized Light and Polarization Vision in Animal Sciences* (Springer, Berlin, 2014).
  138. Stalleicken, J., Labhart, T. & Mouritsen, H. Physiological characterization of the compound eye in monarch butterflies with focus on the dorsal rim area. *J. Comp. Physiol.* **192**, 321–331 (2006).
  139. Kamermans, M. & Hawryshyn, C. Teleost polarization vision: how it might work and what it might be good for. *Phil. Trans. R. Soc. Lond. B* **366**, 742–756 (2011).
  140. Wiltshcko, W., Daum, P., Fergenbauer-Kimmel, A. & Wiltshcko, R. The development of the star compass in garden warblers, *Sylvia borin*. *Ethology* **74**, 285–292 (1987).
  141. Michalik, A., Alert, B., Engels, S., Lefeldt, N. & Mouritsen, H. Star compass learning: how long does it take? *J. Ornithol.* **155**, 225–234 (2014).
  142. Mouritsen, H. & Larsen, O. N. Migrating songbirds tested in computer-controlled Emlen funnels use stellar cues for a time-independent compass. *J. Exp. Biol.* **204**, 3855–3865 (2001).
  143. Alert, B., Michalik, A., Helduser, S., Mouritsen, H. & Güntürkün, O. Perceptual strategies of pigeons to detect a rotational centre—a hint for star compass learning? *PLoS One* **10**, e0119919 (2015).
  144. Dacke, M., Baird, E., Byrne, M., Scholtz, C. H. & Warrant, E. J. Dung beetles use the Milky Way for orientation. *Curr. Biol.* **23**, 298–300 (2013).
  145. Zuffall, F. & Munger, S. *Chemosensory Transduction: The Detection of Odors, Tastes, and Other Chemostimuli* (Academic, London, 2016).
  146. Allison, J. D. & Cardé, R. T. *Pheromone Communication in Moths: Evolution, Behavior, and Application* (Univ. California Press, Oakland, 2016).
  147. Jorge, P. E., Marques, P. A. & Phillips, J. B. Activational effects of odours on avian navigation. *Proc. R. Soc. Lond. B* **277**, 45–49 (2010).
  148. Wallraff, H. G. & Andrae, M. O. Spatial gradients in ratios of atmospheric trace gases: a study stimulated by experiments on bird navigation. *Tellus B Chem. Phys. Meteorol.* **52**, 1138–1157 (2000).
  149. Kullberg, C., Henshaw, I., Jakobsson, S., Johansson, P. & Fransson, T. Fuelling decisions in migratory birds: geomagnetic cues override the seasonal effect. *Proc. R. Soc. Lond. B* **274**, 2145–2151 (2007).
  150. Schmitz, H. & Bleckmann, H. The photomechanic infrared receptor for the detection of forest fires in the beetle *Melanophila acuminata* (Coleoptera: Buprestidae). *J. Comp. Physiol. A* **182**, 647–657 (1998).
  151. Hagstrum, J. T. Infrasonic and the avian navigational map. *J. Exp. Biol.* **203**, 1103–1111 (2000).
  152. Reynolds, A. M., Reynolds, D. R., Sane, S. P., Hu, G. & Chapman, J. W. Orientation in high-flying migrant insects in relation to flows: mechanisms and strategies. *Phil. Trans. R. Soc. Lond. B* **371**, 20150392 (2016).
  153. Sjöberg, S. & Muheim, R. A new view on an old debate: type of cue-conflict manipulation and availability of stars can explain the discrepancies between cue-calibration experiments with migratory songbirds. *Front. Behav. Neurosci.* **10**, 29 (2016).
  154. Åkesson, S. & Bianco, G. Route simulations, compass mechanisms and long-distance migration flights in birds. *J. Comp. Physiol. A* **203**, 475–490 (2017).
  155. Chernetsov, N., Kishkinev, D., Kosarev, V. & Bolshakov, C. V. Not all songbirds calibrate their magnetic compass from twilight cues: a telemetry study. *J. Exp. Biol.* **214**, 2540–2543 (2011).
  156. Hafting, T., Fyhn, M., Molden, S., Moser, M.-B. & Moser, E. I. Microstructure of a spatial map in the entorhinal cortex. *Nature* **436**, 801–806 (2005).
  157. Cheeseman, J. F. et al. Way-finding in displaced clock-shifted bees proves bees use a cognitive map. *Proc. Natl Acad. Sci. USA* **111**, 8949–8954 (2014).

**Acknowledgements** I am grateful to many of the key scientists working in the field of animal navigation and magnetoreception, including all my colleagues and associated members of the proposed collaborative research centre SFB 1372, for inspiration and discussions, for commenting on earlier drafts of this manuscript, and for providing valuable input to various sections of the review. Funding was provided by the Air Force Office of Scientific Research (Air Force Material Command, USAF award no. FA9550-14-1-0095 and FA9550-14-1-0242), the DFG (Graduiertenkolleg 1885, SFB 1372), the ‘Ministerium für Wissenschaft und Kultur’ (Landesgraduiertenkolleg Nano-Energieforschung), and the University of Oldenburg.

**Reviewer information** Nature thanks S. Åkesson, J. Chapman and J. Phillips for their contribution to the peer review of this work.

**Competing interests** The authors declare no competing interests.

#### Additional information

**Reprints and permissions information** is available at <http://www.nature.com/reprints>.

**Correspondence and requests for materials** should be addressed to H.M.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

# Equivalent-accuracy accelerated neural-network training using analogue memory

Stefano Ambrogio<sup>1</sup>, Prithish Narayanan<sup>1</sup>, Hsin-yu Tsai<sup>1</sup>, Robert M. Shelby<sup>1</sup>, Irem Boybat<sup>2,3</sup>, Carmelo di Nolfo<sup>1,3</sup>, Severin Sidler<sup>1,3</sup>, Massimo Giordano<sup>1</sup>, Martina Bodini<sup>1,3</sup>, Nathan C. P. Farinha<sup>1</sup>, Benjamin Killeen<sup>1</sup>, Christina Cheng<sup>1</sup>, Yassine Jaoudi<sup>1</sup> & Geoffrey W. Burr<sup>1\*</sup>

Neural-network training can be slow and energy intensive, owing to the need to transfer the weight data for the network between conventional digital memory chips and processor chips. Analogue non-volatile memory can accelerate the neural-network training algorithm known as backpropagation by performing parallelized multiply-accumulate operations in the analogue domain at the location of the weight data. However, the classification accuracies of such in situ training using non-volatile-memory hardware have generally been less than those of software-based training, owing to insufficient dynamic range and excessive weight-update asymmetry. Here we demonstrate mixed hardware-software neural-network implementations that involve up to 204,900 synapses and that combine long-term storage in phase-change memory, near-linear updates of volatile capacitors and weight-data transfer with ‘polarity inversion’ to cancel out inherent device-to-device variations. We achieve generalization accuracies (on previously unseen data) equivalent to those of software-based training on various commonly used machine-learning test datasets (MNIST, MNIST-backrand, CIFAR-10 and CIFAR-100). The computational energy efficiency of 28,065 billion operations per second per watt and throughput per area of 3.6 trillion operations per second per square millimetre that we calculate for our implementation exceed those of today’s graphical processing units by two orders of magnitude. This work provides a path towards hardware accelerators that are both fast and energy efficient, particularly on fully connected neural-network layers.

Deep neural networks (DNNs) are a family of neuromorphic computing architectures that have recently made substantial advances in difficult machine-learning problems such as image or object recognition, speech recognition and machine language translation<sup>1</sup>. Computation for DNNs includes both training, during which the weights of the network are optimized on a training dataset, and forward inference, during which the already-learned network is used for classification, prediction or other useful tasks on new, previously unseen ‘test’ data.

These networks are highly amenable to computation via large and dense matrix-matrix multiplications that can be highly parallelized. This has led to tremendous opportunities for hardware acceleration by using graphical processing units (GPUs)<sup>1,2</sup>, which in turn enable large networks with commercially interesting levels of performance. Furthermore, DNNs are highly resilient to numerical inaccuracies<sup>3</sup>, especially for forward inference<sup>4</sup>. As a result, reductions in computational precision by using field-programmable gate array (FPGA)<sup>3</sup> and application-specific integrated circuit (ASIC) designs<sup>5,6</sup> offer a path to even higher computational performance and better power efficiency.

Conventional von Neumann hardware is constrained by the time and energy spent moving data back and forth between the memory and the processor (the ‘von Neumann bottleneck’). By contrast, in a non-von Neumann scheme, computing is done at the location of the data, with the strengths of the synaptic connections (the ‘weights’) stored and adjusted directly in memory.

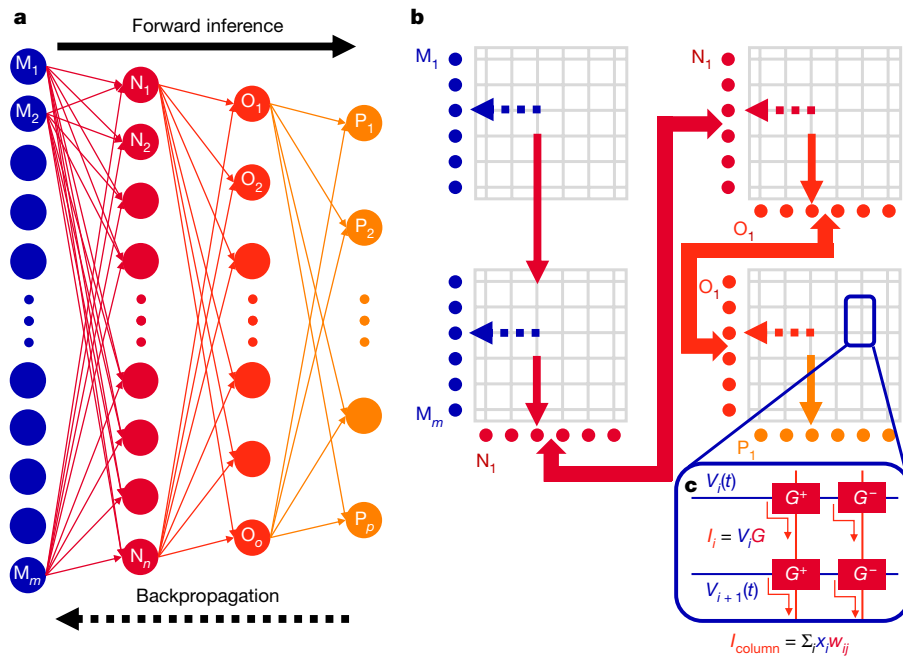
An example of hardware that uses a non-von Neumann scheme is the TrueNorth chip (IBM), a flexible platform for forward inference of large pre-trained DNNs at ultralow power<sup>7,8</sup>. However, for efficient on-chip training, it would be preferable to replace the digital synaptic weights, which are stored in static random access memory (SRAM) arrays on TrueNorth, with high-density analogue devices that encode synaptic weight directly in their conductances. Such analogue systems could

achieve substantial speedup and power reduction for both forward inference and training<sup>9–12</sup>. However, it has not been conclusively proven that such analogue approaches can ‘do the same job’ as current software running on conventional digital hardware, in terms of training DNNs to equivalently high accuracies. There is little point to being faster or more energy-efficient in training a DNN if the resulting classification accuracies are unacceptably low.

Desirable characteristics of analogue devices for training include rapid, low-power programming of multiple analogue levels, dimensional scalability, reasonable retention, high endurance and, most importantly, gradual and symmetric conductance-update characteristics<sup>13</sup>. So far, experimental demonstrations of analogue-memory-based DNN training have suffered from reduced classification accuracies owing to the substantial non-idealities exhibited by existing devices. These demonstrations have featured filamentary resistive RAM (RRAM)<sup>14–16</sup>, non-filamentary resistive RAM<sup>17</sup>, phase-change memory (PCM)<sup>10,11</sup>, conductive-bridging RAM (CBRAM)<sup>18</sup>, ferroelectric RAM<sup>19</sup> and hybrid digital-non-volatile memory (NVM) architectures<sup>20</sup>. Electrochemical devices that offer highly symmetric and gradual conductance update—such as ENODE<sup>21</sup> (electrochemical neuromorphic organic device) and LISTA<sup>22</sup> (lithium-ion synaptic transistor for analogue computing)—have been demonstrated, but not in array configurations and so far only with programming pulse durations that are many orders of magnitude too long.

Here, we introduce a synaptic unit-cell design for analogue-memory-based DNN training, which combines non-volatile PCM with volatile weight storage using conventional complementary metal-oxide-semiconductor (CMOS)-based devices. Using this unit cell, we demonstrate software-equivalent DNN accuracies experimentally for various datasets in fully connected networks. These mixed hardware-software experiments combine PCM hardware arrays with

<sup>1</sup>IBM Research-Almaden, San Jose, CA, USA. <sup>2</sup>IBM Research-Zurich, Rueschlikon, Switzerland. <sup>3</sup>EPFL, Lausanne, Switzerland. \*e-mail: [gwburr@us.ibm.com](mailto:gwburr@us.ibm.com)



**Fig. 1 | Mapping a fully connected neural network onto NVM arrays.** **a, b,** A fully connected four-layer (M, N, O and P) neural network of size  $m$ - $n$ - $o$ - $p$  (**a**) can be mapped to multiple blocks of crossbar arrays surrounded by peripheral neuron circuitry (**b**). For wide layers with many neurons, multiple blocks of arrays are combined. Thin solid arrows in **b** show the column-based current integration during forward inference, which corresponds to the logical flow indicated by solid arrows in **a**.

realistic SPICE (simulation program with integrated circuit emphasis)-based circuit simulations that include full CMOS device variability. We demonstrate neural-network training of the MNIST<sup>23</sup> and MNIST-backrand (https://www.iro.umontreal.ca/~lisa/twiki/bin/view.cgi/Public/MnistVariations) databases—databases of handwritten digits from the National Institute of Standards and Technology (NIST): MNIST, ‘modified’ from original NIST database; MNIST-backrand, MNIST with random background noise added to each image—and transfer learning of CIFAR-10 and CIFAR-100<sup>24</sup> datasets—datasets from the Canadian Institute for Advanced Research with 10 or 100 classes of images. We present power estimates for the device arrays and for the requisite analogue peripheral circuitry, with power projections for DNN training as low as 54 mW for a computational energy efficiency of 28,065 billion operations per second per watt (28,065 GOP s<sup>-1</sup> W<sup>-1</sup>) and throughput per unit area of 3.6 trillion operations per second per square millimetre (3.6 TOP s<sup>-1</sup> mm<sup>-2</sup>). These efficiency and throughput values that we estimate are 280 and 100 times better, respectively, than those achieved using the most-recent GPUs. We believe that these results demonstrate a viable path to low-power hardware acceleration of the training of a wide variety of DNNs using existing analogue memory and computing elements.

### Deep learning using NVM

By mapping synaptic layers onto crossbar array blocks (Fig. 1a), non-von Neumann hardware based on NVM could potentially perform all the multiply-accumulate operations for a fully connected neural-network layer in a single parallelized step, without any motion of weight data. Each network layer maps onto one or more identical array blocks (Fig. 1b), each containing ‘upstream’ (at the end of each row) and ‘downstream’ (at the end of each column) neuron circuitry, interconnected by a flexible block-to-block routing network<sup>25</sup>. Neuron circuitry at the west (south) edges generates these pulses on the basis of the accumulated current from the preceding forward (reverse) pass, which drives the crossbar bit and word lines to appropriate voltages on the basis of the mode of operation and the duration of the neuron pulse<sup>25</sup>.

Similarly, dotted arrows show the row-based current integration during backpropagation; wide arrows in **b** show the routing communication needed to connect the two sets of neuron circuitry corresponding to each neuron layer in **a**. **c**, Details of how neuron excitations encoded as voltage pulses ( $V_i(t)$ , where  $i$  corresponds to the row) get multiplied by weight data encoded as conductances (via Ohm's law,  $I_i = V_i G$ ), after which Kirchhoff's current law is applied along the columns  $j$  ( $I_{\text{column}} = \sum_i x_i w_{ij}$ ).

The analogue signals used during the forward or reverse pass are fixed in amplitude, but vary in duration. This enables standard CMOS buffering approaches to be used when communicating signals from the outputs of one array to the inputs of another array, and makes it possible to avoid the considerable power- and area-overhead of analogue-to-digital converters. Reconfigurable connectivity between arrays enables mapping of different neural-network topologies to the same physical hardware using only a few configuration bits per array.

During forward inference, upstream neurons  $i$  introduce excitations  $x_i$  onto row lines. Ohm's law ( $I = VG$ , where  $I$  is the current,  $V$  is the voltage and  $G$  is the conductance) implements the multiplication between these excitations  $x_i$  and the weight values  $w_{ij}$  that are encoded in the conductance  $G$  of the NVM device. Signed  $w_{ij}$  weights can be encoded into the difference between a pair of conductances,  $G^+ - G^-$  (Fig. 1c). Kirchhoff's current law sums these contributions along each column line and the downstream neurons integrate the overall signal<sup>25</sup>,  $\sum_i x_i w_{ij}$ .

During backpropagation, the parts played during forward inference are reversed, with the downstream neuron  $j$  introducing ‘delta’ values  $\delta_j$  onto the columns and the upstream neuron  $i$  accumulating integration along each row to implement<sup>10,25,26</sup>  $\sum_i \delta_i w_{ij}$ . For weight update, each set of neurons fires either a deterministic<sup>10,27</sup> or stochastic<sup>12</sup> series of pulses, with the number of pulses based on the most recent values of  $x_i$  and  $\delta_j$  passed through those neurons during the forward-inference and backpropagation steps. The overlap of these upstream and downstream pulses adjusts the synaptic weights in parallel across the entire array. This scheme corresponds to a mini-batch of size one, with weight updates occurring with every example.

During the DNN training process, a typical weight will receive many thousands of increase requests and almost the same number of decrease requests. In a software implementation, these opposite-sign contributions cancel each other so only a small fraction of weights change substantially. Unfortunately, nonlinearities or other imperfections can cause conductance changes in NVM-based weights to be

strongly asymmetric<sup>10</sup>, which prevents the opposite-sign contributions from cancelling properly. This update asymmetry is the most important source of poor DNN classification accuracy for NVM-based in-situ neural-network training, followed closely by limited dynamic range (too few conductance steps between the lowest and highest conductance)<sup>11,12,14</sup>.

In the next section, we introduce a unit-cell with both more dynamic range and better update symmetry, thus making software-equivalent training accuracies possible despite the imperfections of existing NVM devices.

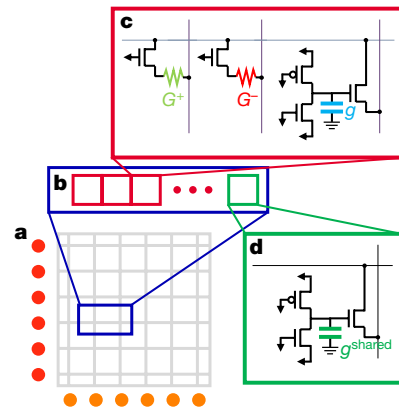
### CMOS + PCM unit cell

We increase the dynamic range by using two pairs of conductance of varying significance (that is, of different numerical importance). This is implemented by applying different scale factors on the read current from the conductance pairs, so that the total read current per synapse is proportional to  $F(G^+ - G^-) + (g^+ - g^-)$ , where  $F$  is a small gain factor (we typically use  $F = 3$ ; Extended Data Fig. 1),  $G^+$  and  $G^-$  are the conductances of the higher-significance conductance pair, and  $g^+$  and  $g^-$  are the conductances of the lower-significance pair. During training, only the lower-significance conductance pair is updated, using an open-loop weight-update procedure<sup>10</sup>. After some number of examples have been trained, weight transfer is initiated and the entire synaptic weight is transferred to the higher-significance conductance pair, after which  $g^+$  and  $g^-$  are programmed to the same conductance ( $g^+ - g^- = 0$ ). Because this is done only periodically (we typically use a transfer interval of 8,000 examples; Extended Data Fig. 1), while the algorithm is busy doing computations on device arrays that correspond to other layers of the neural network, there is sufficient time to perform this transfer process in a closed-loop, iterative fashion, similarly to previously used processes<sup>28–30</sup>, resulting in accurate weight tuning.

A similar ‘periodic carry’ concept involving additional analogue-to-digital converters was shown to improve the expected performance of TaO<sub>x</sub> memristors<sup>22</sup>, as demonstrated by network simulations extrapolated from the measured characteristics of a small number of devices. However, training accuracy was still lower than network simulations based on similar extrapolations from the measured characteristics of a few LISTA<sup>22</sup> or ENODE<sup>21</sup> devices. This is because LISTA and ENODE devices, although exceedingly slow to program (from 6 ms for ENODE<sup>21</sup> to 1 s for LISTA<sup>31</sup>), provide a much more linear conductance update than does TaO<sub>x</sub><sup>21,22</sup>.

PCM offers a slightly more linear conductance update<sup>10</sup> than filamentary RRAM devices<sup>22</sup>, but less linearity than LISTA<sup>22</sup> or ENODE<sup>21</sup>. Our experience using two PCM-based conductance pairs of varying significance was similar to that described for TaO<sub>x</sub><sup>22</sup>: substantial asymmetry in the update of the lower-significance pair, together with a yield of less than 100% and non-ergodic array statistics, which few-device measurements frequently fail to capture, led to improved but not software-equivalent training accuracies.

One analogue memory device that offers extremely linear update characteristics is a CMOS transistor with a capacitor on its gate electrode<sup>32</sup>. The effective conductance of the transistor varies linearly with the voltage on the capacitor, which can be increased (decreased) gradually by briefly connecting current sources to the capacitor node to add (subtract) charge. However, this analogue conductance device is volatile—charge on the capacitor leaks away with a time constant of milliseconds or less. As a result, the neural-network training is tasked not only with improving the weights, but also with maintaining them despite their pervasive and exponential decay. Furthermore, the ideal target of about 1,000 resolvable analogue states<sup>12</sup> and of high linearity calls for as many as 12 transistors in the design of the two current sources<sup>32</sup>, making the unit cell quite large. Finally, fabrication variations can easily cause the charge-addition circuitry in any given unit cell to be more effective than the charge-subtraction circuitry, while in the next unit cell the situation is reversed. This re-introduces substantial asymmetries in the conductance update, which again degrade training accuracy. But by using two such volatile analogue conductance



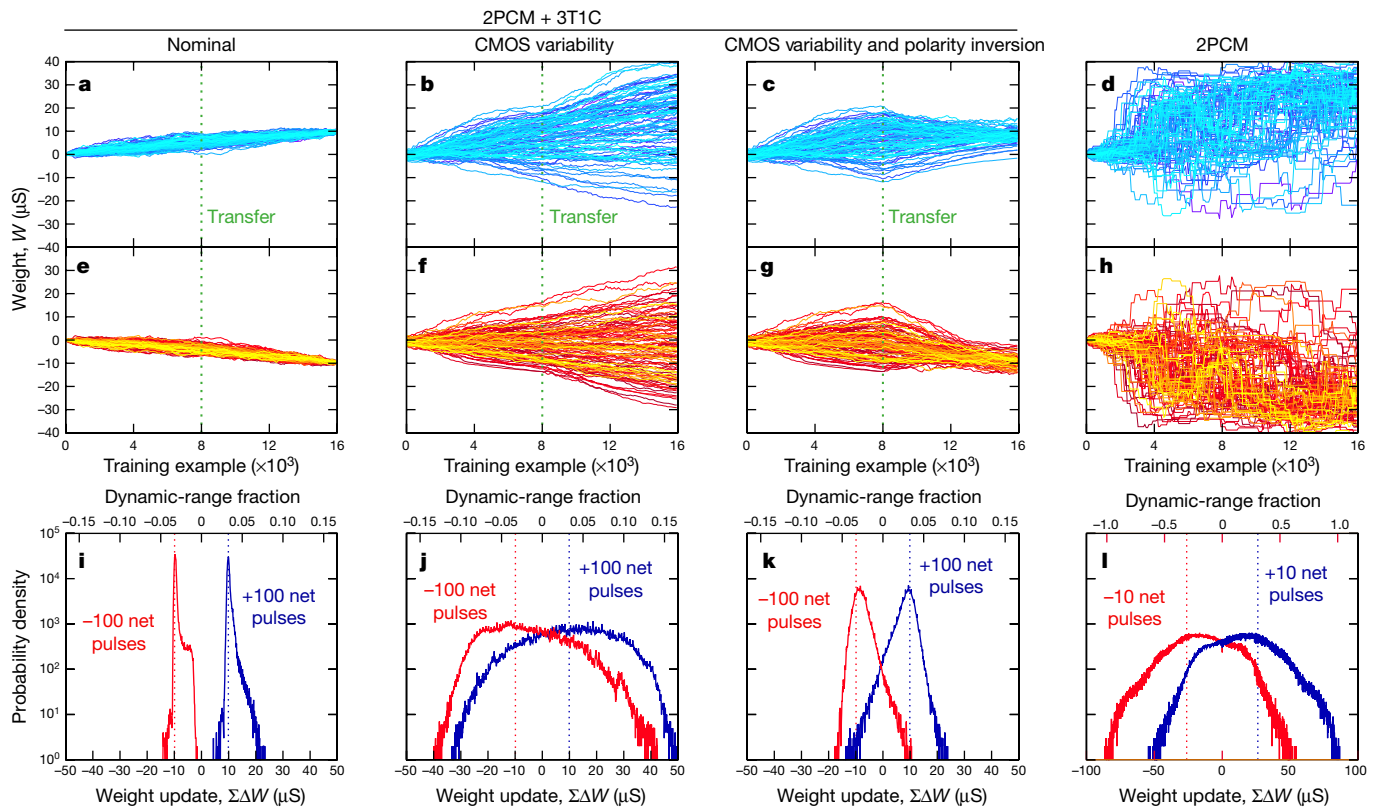
**Fig. 2 | Schematic of an analogue-memory unit cell.** **a, b**, Rows within each array block (**a**) contain both standard (red) and shared (green) unit cells (**b**). **c**, Each standard unit cell contains a higher-significance pair of PCM devices (labelled  $G^+$  and  $G^-$ ) and a volatile analogue conductance (labelled  $g$ ). Horizontal (vertical) arrows indicate which signals are controlled from the same row (column). **d**, Each shared unit cell contains the other volatile analogue conductance (labelled  $g^{\text{shared}}$ ) that completes each lower-significance conductance pair. In our experiments, the peripheral neurons and the  $g$  and  $g^{\text{shared}}$  devices are modelled in software using highly accurate SPICE simulations that include full device variability; the  $G^+$  and  $G^-$  devices are real PCM devices programmed and measured on a 200-mm wafer.

circuit modules as the lower-significance pair (referred to as  $g^+$  and  $g^-$ ) together with a higher-significance non-volatile pair of PCM devices (referred to as  $G^+$  and  $G^-$ ), we combine the benefits of adjusting weights using cells with both a linear and symmetric response, while still retaining the long-term storage offered by the NVM.

Because the dynamic range of the synapse is only partly dependent on the CMOS cell, we can reduce the number of transistors to three. The volatile conductance ( $g$ ) circuit module is composed of the read transistor, a PFET (p-type field-effect transistor) for adding charge and an NFET (n-type field-effect transistor) for subtracting charge (Fig. 2), resulting in a 3T1C (‘3 transistor, 1 capacitor’) circuit module. It can be programmed with shorter pulses and lower asymmetry than can the PCM, leading naturally to rapid training with better characteristics. In contrast to PCM devices, for which the only available gradual programming is conductance increases through partial crystallization, each  $g$  device can be programmed bi-directionally. We can gradually increase the weight by adding some charge to the capacitor through short pulses on the PFET, or decrease the weight similarly through the NFET. Therefore, unlike PCM-based conductance pairs, we do not need to independently program a  $g^-$  device to tune the weights. We do need a reference current to support negative weight contributions, but this device ( $g^{\text{shared}}$ ) can be shared among many unit cells. We adopted one  $g^{\text{shared}}$  device—implemented here with three 3T1C units in parallel to reduce variability—for every 128 unit cells. Not counting these  $g^{\text{shared}}$  components, each dedicated synaptic unit cell contains five transistors, two PCM devices and one capacitor.

During training, only the lower-significance conductance  $g$  is updated bi-directionally until it is time for a weight transfer. The length of a transfer interval is chosen to balance the saturation of the capacitor, the leakage of the capacitor and the costs of doing a transfer (in time, energy and accuracy due to incomplete weight transfer). In an eventual chip implementation, transfer would be performed one column (or row) at a time, with weight information transferred from the  $g - g^{\text{shared}}$  conductance pair to the  $G^+ - G^-$  PCM pair.

However, we still have a problem, owing to unavoidable random variations in CMOS devices (such as local dopant fluctuations). For instance, a 3T1C device with a PFET that is more effective than its NFET will tend to report weight increases at every transfer interval. Because we transferred all of the weight from the  $g - g^{\text{shared}}$  conductance pair, we can choose to invert the effective polarity of this conductance



**Fig. 3 | Simulated response of different unit cells to nearly offsetting weight-update requests. a–l,** Simulation results showing the open-loop conductance-update behaviour of four NVM-based synapses: 2PCM + 3T1C with perfectly nominal CMOS devices (**a, e, i**), 2PCM + 3T1C with real CMOS variability but no polarity inversion (**b, f, j**), 2PCM + 3T1C with real CMOS variability and polarity inversion (**c, g, k**) and 2PCM (**d, h, l**). **a–c,** Effect of the application of 1,000 update pulses, 550 up and 450 down, randomly permuted and distributed over 16,000 examples (other pulses equal to zero). **d,** The effect of 50 pulses, 30 up and 20 down, distributed over 16,000 examples. Different colours in **a–d** show 100 different random permutations. **e–h,** Results of applying 450 up and 550 down pulses for 2PCM + 3T1C (**e–g**) or 20 up and 30 down (**h**). These pulse sequences are based on MNIST results and reveal typical neural-network requests over 16,000 training examples (an equivalent period of two transfer intervals for 2PCM + 3T1C, Extended Data Fig. 2). For **a–c** and **e–g**, weight transfer occurs after 8,000 examples

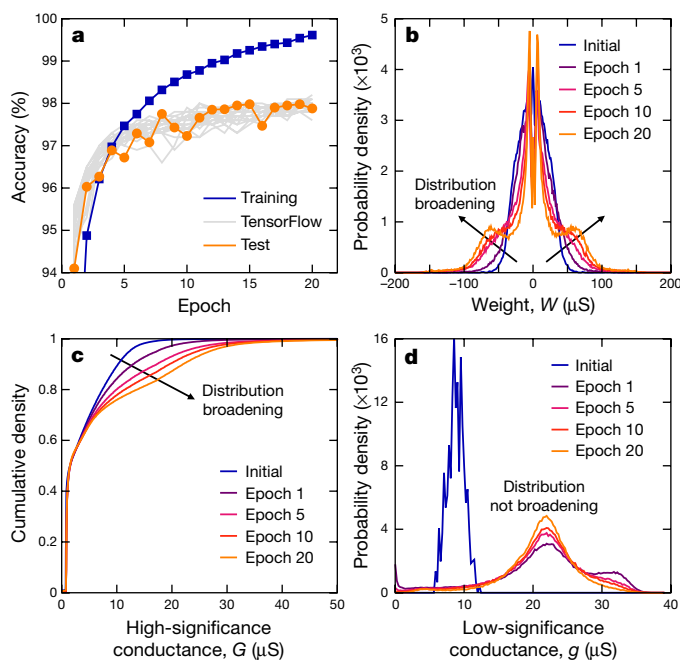
(indicated by the vertical green dashed lines) and at the end; for **d** and **h**, occasional reset<sup>10</sup> is performed every 100 examples. Polarity inversion strongly reduces the effect of CMOS variability. **i–l,** Distribution of the resulting weight update for 30 initial conditions (10,000 random permutations at each) across the entire dynamic range ( $-40\mu\text{S}$  to  $+40\mu\text{S}$ ), not only the initial condition ( $W = 0$ ) shown in **a–h**. The dashed vertical lines show the precise aggregate change in weight that the backpropagation algorithm was ideally seeking with these particular pulse sequences. Tighter distributions correlate well with higher training accuracies (see Figs. 4–6). For **i–k** and **l**, the dynamic-range fraction represents the portion of the dynamic range covered by weights for the application of  $\pm 100$  pulses (from a total of 1,000 pulses) and of  $\pm 10$  pulses (from a total of 50 pulses), respectively. Note that the neural network attempts to compensate for the larger steps in the 2PCM synapses by asking for fewer of them (for example, the optimal learning rate is lower).

pair until the next transfer operation. This involves switching to the equation  $F(G^+ - G^-) - (g - g^{\text{shared}})$  while reading device currents (for both forward inference and backpropagation), using the PFET to add charge when decreasing the weight and the NFET to subtract charge when increasing the weight. After each transfer interval, we invert the polarity used for  $g - g^{\text{shared}}$  during the subsequent training cycle. (See Methods for details.)

In Fig. 3 we compare the operation of four NVM-based synapses when implementing one of the demanding sequences of programming pulses that occur during neural-network training. Matched simulations were performed for a ‘2PCM + 3T1C’ unit cell (which contains two PCM devices, each with its own selection transistor, and one 3T1C circuit module) with perfectly nominal CMOS devices (Fig. 3a, e, i), for 2PCM + 3T1C with full CMOS variability (Fig. 3b, f, j), for 2PCM + 3T1C with CMOS variability and polarity inversion (Fig. 3c, g, k) and, for comparison, two PCM devices without the 3T1C (‘2PCM’) (Fig. 3d, h, l). First, we used MNIST neural-network training simulations to find the correlation between what the network asks for and what the network actually gets (in terms of weight updates). The former is the effective number of weight-change pulses (or ‘net’ pulses) over some time interval, for example, the number of weight-increase requests less the number of weight-decrease requests; the latter is

the actual number of pulses fired and the resulting weight change (Extended Data Fig. 2). We picked one example combination of a net change of  $\pm 100$  pulses obtained by firing exactly 1,000 pulses for the 2PCM + 3T1C cell (Fig. 3a–c, e–g) and of  $\pm 10$  pulses obtained from exactly 50 pulses for 2PCM (Fig. 3d, h). In these simulations, we randomly distribute these 1,000 or 50 pulses across a window of 16,000 training examples (representing two transfer intervals). In Fig. 3i–l, we show weight-change statistics gathered over all possible initial weight conditions.

Each of these pulse sequences should result in exactly the same net number of weight-update pulses and therefore exactly the same weight change (dashed lines in Fig. 3i–l). Although nominal CMOS 3T1C devices provide clearly separated weight increases and decreases (Fig. 3a, e), CMOS variability strongly broadens the weight change (Fig. 3b, f). This occurs because a 3T1C cell in which the PFET is stronger than its NFET always favours weight increases over every transfer interval. By forcing that strong PFET to be responsible for weight decreases in every second transfer interval, a longer-term balance can be restored. A similar argument holds for the cells with a strong NFET. Figure 3c, g illustrates how, across multiple transfer intervals, ‘transfer with polarity inversion’ cancels out these undesired weight changes that are induced by fixed device asymmetry.



**Fig. 4 | Mixed hardware-software results on the MNIST dataset.**

**a**, Training (blue) and test (orange) accuracies for our mixed hardware-software experiment, which combines hardware-based PCM devices and SPICE-modelled 3T1C devices with full CMOS variability, on the MNIST dataset closely match those achieved for the same size network using TensorFlow (grey). The TensorFlow curves correspond to ten different initial network conditions and sequencing of training images, illustrating the modest run-to-run deviations that are inherent to neural-network training. **b**, **c**, The distribution of weights for the initial state and epochs 1, 5, 10 and 20 (**b**), and the cumulative distributions of all array conductances ( $G^+$  and  $G^-$  together; **c**). Half of the  $G^+$  ( $G^-$ ) values are almost zero, corresponding to negative (positive) weights. **d**, Initial  $g$  distribution and successive distributions just before weight transfer to the hardware-based PCM devices.

In comparison to the 3T1C-based synapses, 2PCM synapses offer much less dynamic range.

Although Fig. 3 demonstrates that both the 2PCM + 3T1C concept and transfer with polarity inversion should lead to better training accuracies, it is extremely difficult to specify how tight these distributions should be to achieve a given training accuracy. It is therefore critical to perform actual neural-network training experiments, which we implement below with real PCM devices. This enables us to demonstrate that transfer with polarity inversion results in high training accuracy even in the presence of substantial variability of the CMOS device.

### Mixed hardware-software implementation

We conducted mixed hardware and SPICE model experiments using PCM device arrays as hardware (identical to those used previously<sup>10</sup>) and realistic software-based CMOS device models, including highly accurate modelling of the variability of the CMOS device. Training within a transfer interval was performed in software, including forward propagation, backpropagation and weight update of the modelled 3T1C devices. Each software synapse contains the measured conductances from two PCM devices, the instantaneous voltage of the software-modelled capacitor and the indices for four SPICE models (one for charge addition, one for charge subtraction, one for the read transistor and one for the aggregate charge leakage) that encapsulate full CMOS variability. (See Methods for details; in Extended Data Fig. 1 we compare a fully hardware implementation to the mixed software-hardware experiment demonstrated here.)

To speed up training, we applied an ‘example triage’ method after each forward propagation to decide whether to perform the back-propagation step. If the network was already ‘good’ at classifying the

example, then training was skipped. This helped to reduce the number of examples trained and markedly decreased the wall-clock time of the experiment. By the end of training, approximately 78% of the training examples in the MNIST, 5% in the MNIST-backrand, 47% in the CIFAR-10 and 13% in the CIFAR-100 datasets were being skipped.

### Results

We performed a series of experiments on different benchmark datasets to compare the performance of networks based on our unit cell against the accuracies obtained with TensorFlow, a widely used machine-learning toolkit (<https://www.tensorflow.org>). Our goal was to achieve comparable test accuracy against software baselines that use exactly the same network size as our experiment, but take advantage of software techniques such as unbounded rectified linear unit activation function and cross-entropy training, and optimizers such as AdaGrad, ideal momentum and ADAM<sup>33</sup>. All experiments described below incorporate the full extent of variability and non-ideality seen in the PCM devices and the SPICE-simulated 3T1C cells.

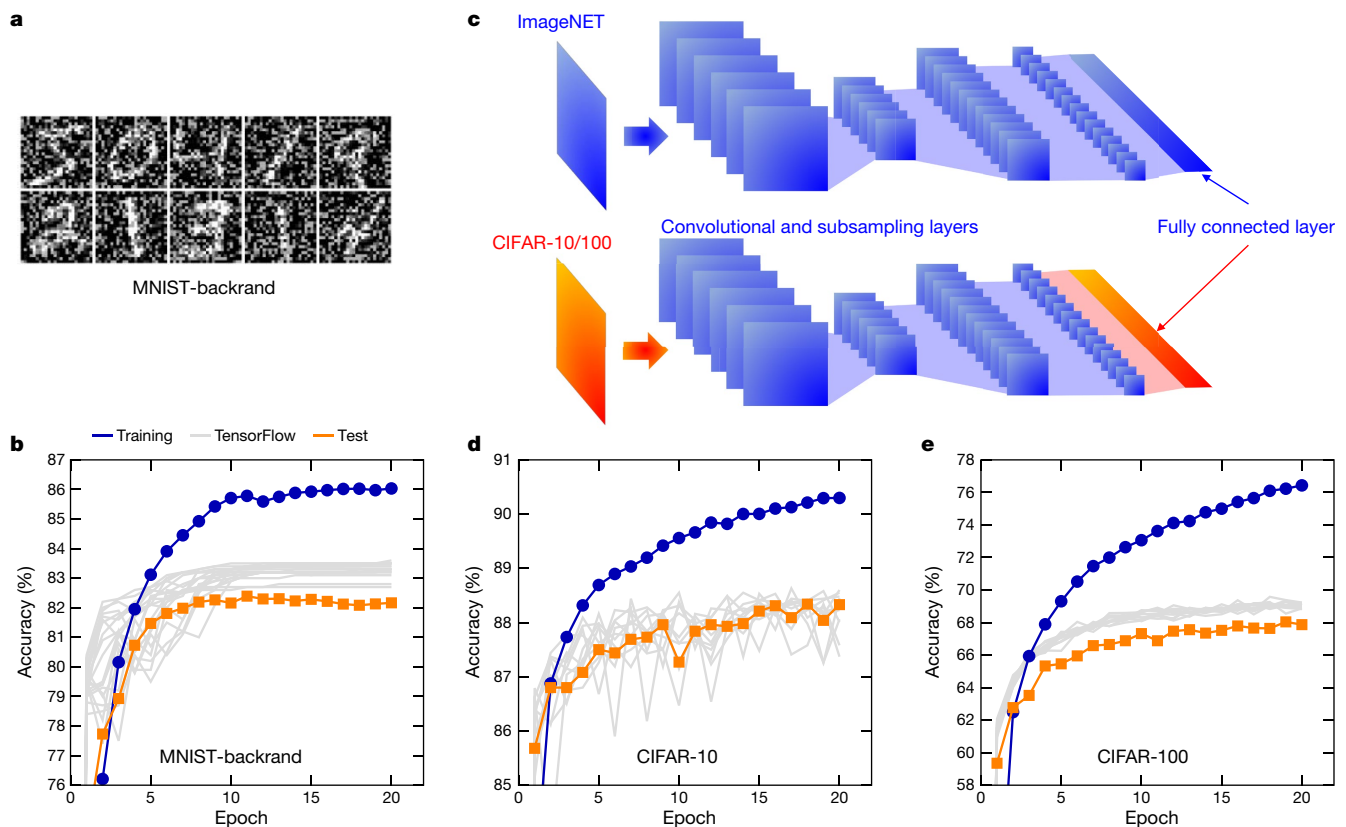
We first tested our NVM-based unit cell on the MNIST dataset<sup>23</sup>, which is composed of 60,000 training and 10,000 test images of handwritten digits, cropped to  $22 \times 24$  pixels. We trained a four-layer network with 528 neurons (plus 1 bias neuron) in the first layer, 250 (plus 1 bias) in the second, 125 (plus 1 bias) in the third and 10 in the fourth (denoted as 528-250-125-10), involving 164,885 weights and therefore 329,770 PCMs. Training and test accuracies over 20 epochs are shown in Fig. 4a, compared against the fully software simulations performed in TensorFlow for the same network. Experimental results closely match the mean software test accuracy of 97.94%.

To track the behaviour of weights, in Fig. 4b we show the evolution of the weight distribution, combining contributions from higher-significance PCM conductances and lower-significance SPICE-simulated  $g - g^{\text{shared}}$  conductances. We also show the PCM conductance distributions during training (Fig. 4c) and the distributions for  $g$  during training at the instant before a weight-transfer operation (Fig. 4d). As training proceeds, weight distributions broaden, as reflected in the broader cumulative distributions of the  $G$  values (Fig. 4c). However, distributions of  $g$  at later epochs do not differ from each other (Fig. 4d), meaning that weight evolution is implemented mainly by the PCM devices. The role of  $g$  is then to support the proper tuning of PCM rather than to encode important long-term information.

The MNIST-backrand dataset poses a much more difficult recognition problem than does MNIST, by adding uniform random noise to each handwritten-digit image (Fig. 5a) and by providing fewer training examples (12,000) but requiring generalization across many more test examples (50,000). We trained on a network comprising 784-180-125-10 neurons. Our NVM-based experimental accuracy of 82.13% is only slightly below the TensorFlow accuracy of 83.3% (Fig. 5b).

State-of-the-art image classification systems use a combination of convolution layers that act as feature extractors with one or more fully connected classification layers. Although our work on efficiently mapping convolution networks to analogue array architectures is ongoing, a well-known approach (transfer learning<sup>34</sup>) repurposes convolution layers that are pre-trained on one dataset for new datasets, by re-training only the last fully connected classification layers. Here, we use the weights from the Google Inception-v3 network (<https://github.com/tensorflow/models/tree/master/research/inception>) (more than 70 layers) trained on ImageNET<sup>35</sup> and re-train a single software-hardware fully connected layer on either the CIFAR-10 or CIFAR-100 dataset.

We used an image re-training script ([https://www.tensorflow.org/tutorials/image\\_retraining](https://www.tensorflow.org/tutorials/image_retraining)) to rescale the  $32 \times 32$  CIFAR images to the  $299 \times 299$  ImageNET input size and to compute the 2,048 neuron excitations at the input of the fully connected layer of Inception-v3 using forward inference. These vectors were then used to train a 2,048-10 fully connected network with two neuron layers with CIFAR-10 labels and a 2,048-100 network with CIFAR-100 labels. Our experimental results are again compared against software-based (TensorFlow) training of these two networks.



**Fig. 5 | Mixed hardware-software results on the MNIST-backrand and CIFAR-10/100 datasets.** **a, b,** When trained on the MNIST-backrand dataset, which consists of handwritten digits plus uniform random noise<sup>23</sup> (**a**), our mixed hardware-software experiment shows similar test accuracies to fully software-based training using TensorFlow (**b**). **c,** For our transfer-learning experiments, the weights from the

Inception-v3 network as originally trained on ImageNET (top) are used to learn CIFAR-10/100 images by re-training only the last fully connected layer (bottom). **d, e,** The experimental test accuracies for CIFAR-10 (**d**) and CIFAR-100 (**e**) closely match the expected results from software-based (TensorFlow) training. See Extended Data Fig. 3 for corresponding experimental cumulative distribution functions.

Training and test accuracies are shown in Fig. 5d for CIFAR-10 (2048-10 neurons, 20,490 weights) and in Fig. 5e for CIFAR-100 (2048-100 neurons, 204,900 weights, corresponding to 409,800 PCM devices). Accuracy is equivalent for CIFAR-10, whereas CIFAR-100 shows a small difference of just over 1% (Extended Data Fig. 3 shows weight and PCM conductance distributions for MNIST-backrand and CIFAR-10/100 experiments). In Fig. 6a we summarize all of our experimental results against the expectations of software-based (TensorFlow) training.

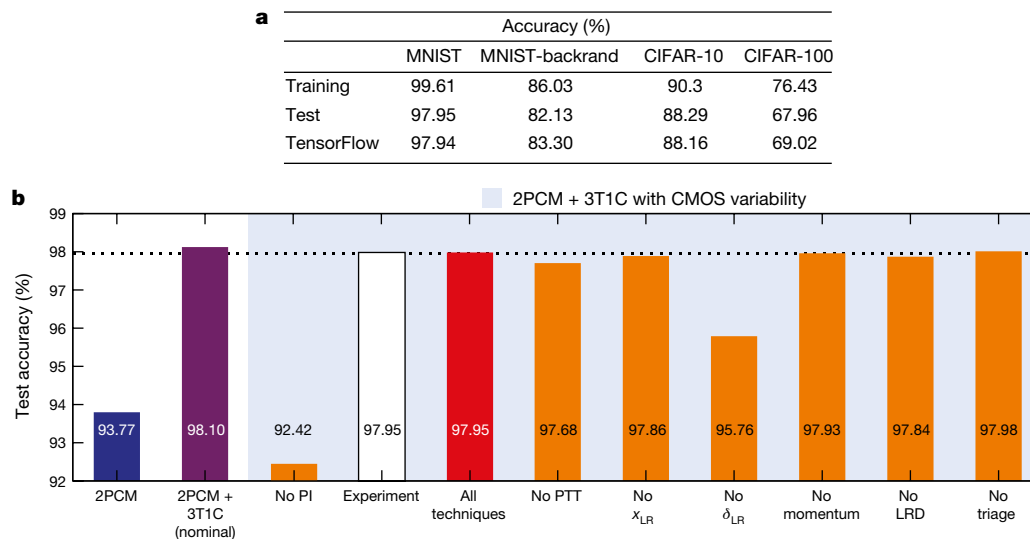
In Fig. 6b we provide a detailed comparison of the MNIST test accuracy after 20 epochs for different unit cells according to a matched simulator that mimics our PCM hardware<sup>10</sup>, as compared to our final mixed hardware-software experiment (Fig. 4 and unfilled bar in Fig. 6b). Adding the 3T1C devices to the flawed 2PCM devices improves the accuracy (purple bar in Fig. 6b) until we consider the strong effect of real CMOS variability (left-most orange bar). As predicted by Fig. 3, polarity inversion greatly reduces the undesirable effects of CMOS variability (red bar in Fig. 6b). We also use this matched simulator to determine how training would have proceeded if we had left out only one of the techniques that we used in the full experiment (orange bars; see also Methods and Extended Data Fig. 4).

Finally, we estimated the average power required to fully process a single MNIST example on the 528-250-125-10 network. We include dissipated power for one full forward and reverse pass, the associated weight updates and a prorated number of occasional reset events and/or transfers per example, on the basis of the activity levels observed within our experiments and on circuit simulations of a full mixed-signal design in 90-nm node. We do not attempt to include the input-output power associated with bringing example and label data onto a chip.

Separate estimates were performed for the 2PCM and 2PCM + 3T1C designs, including power consumption in the three crossbar arrays, and in the peripheral analogue and digital circuitry (see Methods for details).

Under the assumption that an MNIST example can be processed in 240 ns, the average power consumption for the 2PCM + 3T1C design was calculated to be 54 mW, compared to 22 mW for the PCM design. The increase in power of a factor of 2.5 is primarily due to the large forward- and reverse-propagate currents in the 3T1C array, wherein each cell could contribute several microamps of current. To compare to a modern GPU, the training of a single example on our small network performs 362,405 multiply-accumulate operations and requires 12.9 nJ of power for a computational energy efficiency of  $28,065 \text{ GOP s}^{-1} \text{ W}^{-1}$ . At the specified processing time of 240 ns, and extrapolating our 90-nm circuit designs to 14 nm, the throughput per unit area is  $3,582 \text{ GOP s}^{-1} \text{ mm}^{-2}$ . For comparison, a Tesla V100 GPU offers<sup>36</sup> 30.0 trillion floating-point operations per second (30.0 TFLOPs) of 16-bit floating-point numbers in a footprint of 300 W and 815 mm<sup>2</sup>, or  $100 \text{ GOP s}^{-1} \text{ W}^{-1}$  and  $37 \text{ GOP s}^{-1} \text{ mm}^{-2}$ . Therefore, our analogue NVM-based approach could potentially provide more than two orders of magnitude in energy efficiency while accelerating the backpropagation algorithm for fully connected layers by nearly two orders of magnitude.

We have established that our approach can also deliver software-equivalent training accuracies, despite the imperfections of existing analogue memory devices. The next steps will be to demonstrate this same software equivalence on larger networks that require large fully connected layers—such as recurrently connected long short-term memory<sup>37</sup> and gated recurrent networks<sup>38</sup>—and to design, implement and refine these analogue techniques on prototype NVM-based hardware accelerators.



**Fig. 6 | Accuracy comparison and effect of different techniques.**

**a**, Training and test accuracies from our mixed hardware–software experiments on different datasets are compared to expected results from purely software-based (TensorFlow) training. **b**, Using a matched simulator that mimics how our PCM hardware trains on the MNIST dataset, we evaluate the relative effect of the various imperfections present in our experiments and of each mitigation technique that we used within our final experiment by removing that technique during simulated training (orange bars). The matched simulator when no techniques are removed (red bar) closely matches our experimental results (unfilled

bar and dotted horizontal line). The accuracy with only our flawed PCM devices (2PCM) would be insufficient (blue bar); although nominal 3T1C devices could improve this greatly (purple bar), the effect of CMOS variability (bars on light-blue background) is severe without polarity inversion (PI; left-most orange bar). Additional techniques that are important include post-transfer tuning (PTT) of  $g$  after weight transfer and modulation of upstream ( $x_{LR}$ ) and downstream ( $\delta_{LR}$ ) weight-update pulses (see Methods and Extended Data Fig. 4). Other techniques, which are less relevant include momentum, learning-rate decay (LRD) and triage.

We anticipate that the 3T1C cell proposed here will eventually be phased out in favour of a more compact, modestly linear NVM, and that these NVM devices will be stacked above each other in the metal back-end using non-silicon access devices<sup>39</sup>. Although the periodic carry concept<sup>22</sup> has already relaxed device requirements for effective dynamic range, our approach for multiple significant device pairs further relaxes and differentiates these requirements. Our approach calls for a new type of analogue memory device that can provide high linearity and endurance, but that does not need long retention, a huge resistance window or even tightly constrained device-to-device variability. For the higher-significance conductance-pair, we need an NVM device that offers good retention and fast, low-power programming for high-precision closed-loop tuning, but we do not need to impose requirements on ultra-linear conductance update, high endurance or wide dynamic range as well.

Therefore, there remains a strong incentive to develop compact NVM devices that are capable of gentle, symmetric conductance change<sup>12,13</sup>. Even in such an advanced design, we anticipate that the technique used here (and proposed independently elsewhere<sup>22</sup>) of using multiple conductances of varying significance will be necessary to synthesize a synapse with high dynamic range from individual devices of lower dynamic range, and that polarity inversion on transfer (introduced here to suppress the highly undesirable effects of fixed device asymmetries) will be essential.

## Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0180-5>.

Received: 26 January 2018; Accepted: 29 March 2018;

Published online 6 June 2018.

1. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
2. Coates, A. et al. Deep learning with COTS HPC systems. In *Proc. 30th International Conference on Machine Learning* 1337–1345 (Association for Computing Machinery, 2013).

3. Gupta, S., Agrawal, A., Gopalakrishnan, K. & Narayanan, P. Deep learning with limited numerical precision. In *Proc. 30th International Conference on Machine Learning* 1737–1746 (Association for Computing Machinery, 2015).
4. Merolla, P., Appuswamy, R., Arthur, J., Esser, S. K. & Modha, D. Deep neural networks are robust to weight binarization and other non-linear distortions. Preprint at <https://arxiv.org/abs/1606.01981> (2016).
5. Nurvitadhi, E. et al. Can FPGAs beat GPUs in accelerating next-generation deep neural networks? In *Proc. 2017 ACM/SIGSA International Symposium of Field-Programmable Gate Arrays* 5–14 (Association for Computing Machinery, 2017).
6. Jouppi, N. P. et al. In-datacenter performance analysis of a tensor processing unit. In *Proc. 2017 International Symposium on Computer Architecture* 1–12 (Association for Computing Machinery, 2017).
7. Merolla, P. A. et al. A million spiking-neuron integrated circuit with a scalable communication network and interface. *Science* **345**, 668–673 (2014).
8. Esser, S. K. et al. Convolutional networks for fast, energy-efficient neuromorphic computing. *Proc. Natl Acad. Sci. USA* **113**, 11441–11446 (2016).
9. Morie, T. & Amemiya, Y. An all-analog expandable neural network LSI with on-chip backpropagation learning. *IEEE J. Solid-State Circuits* **29**, 1086–1093 (1994).
10. Burr, G. W. et al. Experimental demonstration and tolerancing of a large-scale neural network (165,000 synapses), using phase-change memory as the synaptic weight element. In *2014 IEEE International Electron Devices Meeting* T29.5 (IEEE, 2014).
11. Burr, G. W. et al. Experimental demonstration and tolerancing of a large-scale neural network (165,000 synapses), using phase-change memory as the synaptic weight element. *IEEE Trans. Electron Dev.* **62**, 3498–3507 (2015).
12. Gokmen, T. & Vlasov, Y. Acceleration of deep neural network training with resistive cross-point devices: design considerations. *Front. Neurosci.* **10**, 333 (2016).
13. Burr, G. W. et al. Neuromorphic computing using non-volatile memory. *Adv. Physics X* **2**, 89–124 (2017).
14. Yu, S. et al. Scaling-up resistive synaptic arrays for neuro-inspired architecture: challenges and prospect. In *2015 IEEE International Electron Devices Meeting* 17.3 (IEEE, 2015).
15. Gao, L. et al. Fully parallel write/read in resistive synaptic array for accelerating on-chip learning. *Nanotechnology* **26**, 455204 (2015).
16. Prezioso, M. et al. Training and operation of an integrated neuromorphic network based on metal-oxide memristors. *Nature* **521**, 61–64 (2015).
17. Jang, J.-W., Park, S., Burr, G. W., Hwang, H. & Jeong, Y.-H. Optimization of conductance change in  $\text{Pr}_{1-x}\text{Ca}_x\text{MnO}_3$ -based synaptic devices for neuromorphic systems. *IEEE Electron Device Lett.* **36**, 457–459 (2015).
18. Jeong, Y. J., Kim, S. & Lu, W. D. Utilizing multiple state variables to improve the dynamic range of analog switching in a memristor. *Appl. Phys. Lett.* **107**, 173105 (2015).
19. Kaneko, Y., Nishitani, Y. & Ueda, M. Ferroelectric artificial synapses for recognition of a multishaded image. *IEEE Trans. Electron Dev.* **61**, 2827–2833 (2014).

20. Nandakumar, S. R. et al. Mixed-precision training of deep neural networks using computational memory. Preprint at <https://arxiv.org/abs/1712.01192> (2017).
  21. van de Burgt, Y. et al. A non-volatile organic electrochemical device as a low-voltage artificial synapse for neuromorphic computing. *Nat. Mater.* **16**, 414–418 (2017).
  22. Agarwal, S. et al. Achieving ideal accuracies in analog neuromorphic computing using periodic carry. In *2017 Symposium on VLSI Technology* T13.2 (IEEE, 2017).
  23. Lecun, Y., Bottou, L., Bengio, Y. & Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **86**, 2278–2324 (1998).
  24. Krizhevsky, A. *Learning Multiple Layers of Features From Tiny Images*. Ch. 3, <https://www.cs.toronto.edu/~kriz/cifar.html> (2009).
  25. Narayanan, P. et al. Towards on-chip acceleration of the backpropagation algorithm using non-volatile memory. *IBM J. Res. Develop.* **61**, 11 (2017).
  26. Rumelhart, D. E., Hinton, G. E. & Williams, R. J. Learning representations by backpropagating errors. *Nature* **323**, 533–536 (1986).
  27. Xu, Z. et al. Parallel programming of resistive cross-point array for synaptic plasticity. *Procedia Comput. Sci.* **41**, 126–133 (2014).
  28. Papandreou, N. et al. Programming algorithms for multilevel phase-change memory. In *2011 IEEE International Symposium on Circuits and Systems* 329–332 (IEEE, 2011).
  29. Alibart, F., Gao, L., Hoskins, B. D. & Strukov, D. B. High-precision tuning of state for memristive devices by adaptable variation-tolerant algorithm. *Nanotechnology* **23**, 075201 (2012).
  30. Hu, M. et al. Dot-product engine for neuromorphic computing: programming 1T1M crossbar to accelerate matrix-vector multiplication. In *Proc. 53rd Annual Design Automation Conference* 19 (Association for Computing Machinery, 2016).
  31. Fuller, E. J. et al. Li-ion synaptic transistor for low power analog computing. *Adv. Mater.* **29**, 1604310 (2017).
  32. Kim, S., Gokmen, T., Lee, H.-M. & Haensch, W. E. Analog CMOS-based resistive processing unit for deep neural network training. In *2017 IEEE 60th International Midwest Symposium on Circuits and Systems* 422–425 (IEEE, 2017).
  33. Goodfellow, I., Bengio, Y. & Courville, A. *Deep Learning* Ch. 8 (MIT Press, 2016).
  34. Donahue, J. et al. DeCAF: a deep convolutional activation feature for generic visual recognition. In *Proc. 31st International Conference on Machine Learning* 647–655 (Association for Computing Machinery, 2014).
  35. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. & Wojna, Z. Rethinking the Inception architecture for computer vision. Preprint at <https://arxiv.org/abs/1512.00567> (2015).
  36. Mujtaba, H. Nvidia Volta GV100 12nm FinFET GPU detailed – Tesla V100 specifications include 21 billion transistors, 5120 CUDA cores, 16 GB HBM2 with 900 GB/s bandwidth. *Wccfttech* <https://wccfttech.com/nvidia-volta-gv100-gpu-tesla-v100-architecture-specifications-deep-dive/> (2017).
  37. Hochreiter, S. & Schmidhuber, J. Long short-term memory. *Neural Comput.* **9**, 1735–1780 (1997).
  38. Cho, K., van Merriënboer, B., Bahdanau, D. & Bengio, Y. On the properties of neural machine translation: Encoder-decoder approaches. Preprint at <https://arxiv.org/abs/1409.1259> (2014).
  39. Burr, G. W. et al. Access devices for 3D crosspoint memory. *J. Vac. Sci. Technol. B* **32**, 040802 (2014).
- Acknowledgements** We acknowledge management support from B. Kurdi, C. Lam, W. Wilcke, S. Narayan, T. C. Chen, W. Haensch, R. Divakaruni, J. Welsner and D. Gil, and discussions with P. Solomon, S. Kim, A. Sebastian, K. Hosokawa and S. C. Lewis. This work was performed as part of the 'Neuromorphic Devices & Architectures' project under the auspices of the IBM Research Frontiers Institute (<https://www.research.ibm.com/frontiers>). We acknowledge advice and support from H. Riel, S. Gowda, D. Maynard and the member companies of the IBM RFI.
- Reviewer information** Nature thanks G. C. Adam, R. Legenstein and the other anonymous reviewer(s) for their contribution to the peer review of this work.
- Author contributions** G.W.B. developed the multiple-conductances-of-varying-significance and polarity-inversion techniques; P.N. and G.W.B. designed the 3T1C unit cell; G.W.B., R.M.S., C.d.N., I.B. and P.N. developed the neural-network simulation software; R.M.S., I.B., C.d.N., S.S., M.B., N.C.P.F. and S.A. used the simulator to develop insights key to the success of the experiment; G.W.B. designed and S.A. extended the experimental apparatus; S.A. performed the experiments; H.T. designed the transfer learning experiment and performed the TensorFlow software training; P.N., G.W.B. and S.A. developed the SPICE modelling approach; P.N. performed the power analysis; M.G., S.A. and G.W.B. developed the triage approach used in the experiment; and all authors contributed to the writing and editing of the manuscript.
- Competing interests** The authors declare no competing interests.
- Additional information**
- Extended data** is available for this paper at <https://doi.org/10.1038/s41586-018-0180-5>.
- Reprints and permissions information** is available at <http://www.nature.com/reprints>.
- Correspondence and requests for materials** should be addressed to G.W.B.
- Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## METHODS

**Analogue-memory-based neural networks.** The crossbar-memory approach (Fig. 1) is most efficient for fully connected layers. Training of convolutional layers<sup>23</sup>, in which many neurons share and re-use a small set of weight kernels, is much less straightforward to implement using these crossbar-array techniques. Such networks also pose reduced memory-size and memory-bandwidth demands on conventional digital hardware. Given the reduced memory-bandwidth demands, the recent trend in convolutional networks has been to replace fully connected layers with convolutional layers. Interestingly, the fully parallel computation offered by an analogue crossbar-memory approach would call for exactly the opposite strategy—restricting the number of convolutional layers to the absolute minimum necessary and using transfer learning when possible, especially for the first few layers where the depth dimension is much smaller than the input size. By contrast, for forward inference only, the analogue-memory approach can deliver high speed ( $\text{TOP s}^{-1}$ ) and good energy efficiency ( $\text{TOP s}^{-1} \text{W}^{-1}$ ) for convolutional layers, by replicating kernel weights across multiple large arrays (with some reduction in  $\text{TOP s}^{-1} \text{mm}^{-2}$ ).

**Example triage.** Example triage is a technique for accelerating any supervised learning algorithm of DNNs by skipping over training examples that the network already ‘knows’ relatively well, as quantified by the ‘safety margin’. The safety margin is the difference between the excitation level of the correct output neuron and the highest excitation level of the other output neurons (Extended Data Fig. 5). A positive safety margin denotes a correct classification and a negative safety margin denotes an incorrect classification. After forward inference, each example is assigned a focus probability, depending on the safety margin and the stage of training (Extended Data Fig. 5). Any training example with a safety margin below ‘acceptable safety margin’ is included for training with 100% probability. As the network trains, example triage selects fewer images for backpropagation, as shown by cumulative distributions of the safety margin (Extended Data Fig. 6). Thus, the acceleration boost is large when training accuracy is high (our 528–250–125–10 network trained on MNIST), but smaller when training accuracy is low (our 784–180–125–10 network trained on MNIST-backrand).

**Neural-network implementation.** Although we use a logistic squashing function at each neuron layer in these experiments, we have previously shown that piecewise-linear or ‘hard’ logistic functions are effectively equivalent<sup>40</sup>. In the final layer, we train against the raw difference between ground truth and the output activation, but omit the derivative of the logistic function. This secures the primary benefit of training against cross-entropy loss without the need for expensive softmax exponential functions to explicitly estimate cross-entropy. In our experiments and simulations, no discrepancy in test accuracy could be traced to this minimization choice. Our multiple-conductances-of-varying-significance concept, although similar in high-level description to previous work<sup>22</sup>, was developed independently and therefore provides some very different characteristics. Because we weight the read currents in the analogue rather than the digital domain, we do not need to increase the number of analogue-to-digital converters. (Compared to the array itself, analogue-to-digital converters are particularly power- and area-inefficient.) Our circuitry for integrating charge and for transferring excitation data to the next layer is completely independent of the number of conductances encoding the weights. Finally, as mentioned in the main text, by introducing the concept of using a different device for each tier of significance, we can relax the device constraints on these memory elements much further than with the previous approach<sup>22</sup>.

**Hardware set-up.** A single array diagnostic monitor on a 200-mm wafer was contacted with a 94-pin probe card mounted on a Cascade microprobe station, enabling access to a  $512 \times 1,024$  1T–1PCM device array (180-nm technology). A 10-bit digital bus selects one of 1,024 rows; a second 9-bit bus selects which of the 512 active columns connects to a single master bitline connected to the single read/write head. Both the probe station and the Nextest Magnum 2EV tester were controlled by a single computer running custom neural-network software, the tester code for accessing the PCM arrays through the Magnum and the Matlab interface code for connecting these two software components. Each experiment uses a single large contiguous rectangular block of devices from the entire array, within which all synapses were assigned in order. No remapping to avoid dead or defective devices was performed. Typically, 0.01% of the PCM devices were stuck ‘on’ (persistently remained in conductance states higher than 50  $\mu\text{S}$ ) and 1% of the devices were stuck ‘off’ (persistently remained in conductance states lower than 0.5  $\mu\text{S}$ ). In this mixed hardware–software experiment, we serially read the array, load the values into the software and perform the summation in software. However, in an eventual chip implementation, the vector–matrix multiplication would be completely parallel within each array core.

Before beginning an experiment, weights were initialized to a uniform distribution of raw conductances between 1  $\mu\text{S}$  and 12  $\mu\text{S}$ . The neural-network software reads the PCM array and maps measured conductances into weights within software memory by multiplying the conductance differences by  $F=3.0$ . Roughly 8,000 images are then trained fully in software, including reads and writes of each

3T1C device (from 0 to 7 programming pulses of 300 ps each) according to its set of three SPICE variability models: one for charge addition, one for charge subtraction and one for voltage-to-conductance mapping.

The software keeps track of incurred time (80 ns per forward propagation; 160 ns per backpropagation/weight-update step), decays each capacitor voltage (5.16-ms time constant) and triggers a transfer event after every 1.92 ms of incurred time. The total weight information trained onto each software-based 3T1C device is used to compute a new target weight for the hardware-based PCM devices. After an initial reset pulse, the hardware-based PCM conductances are tuned using an iterative sequence of partial-set pulses varying between 13.6 ns and 4.4  $\mu\text{s}$ . Given the serial hardware read/write, the wall-clock time between two transfer events is 7–20 min, depending on the size of the network and the corresponding number of PCM devices used. PCM drift<sup>41</sup> is present, and contributes to inaccurate PCM conductance tuning, but on a timescale different from that of the internal clock of the software training. After hardware tuning is complete, all hardware-based PCM conductances are measured and reported to the software, which then completes the transfer process. The operational polarity of the software-based 3T1C conductance  $g$  is inverted. Because the PCM programming operation is inherently noisy, the transferred weight is different from the expected weight. To minimize this error, post-transfer tuning is performed: each  $g$  device is programmed using a small number of 300-ps pulses to correct any residual weight error left after hardware-based PCM tuning.

**Weight transfer.** Weight transfer converts large differences in the lower-significance conductance pair into smaller changes in the higher-significance conductance-pair, allowing the network to protect larger weights effectively from the effects of nonlinearity and asymmetry. Any weight residue that cannot be transferred completely to the higher-significance conductance pair can be restored to the lower-significance conductance pair. An eventual chip implementation could implement weight transfer on a column-by-column basis, implemented while the chip is executing a later (or earlier) portion of the network. However, because the hardware described above is most efficient when performing the same operation sequentially (a read comparison or a write operation) on all devices within the array, here weight transfer is performed on the entire array in the same PCM tuning process.

Given the target PCM weights (supplied by the neural-network software in units of raw PCM conductance, after dividing by  $F=3.0$ ), the closed-loop PCM conductance-tuning operation begins with a measurement of all conductances. Any weights already within  $\pm 240$  nS of their targets (typically 2% of the array) are removed from consideration. All remaining devices are first reset with three high-amplitude 500-ns pulses. PCM devices corresponding to  $G^+$  (for positive weights) or  $G^-$  (for negative weights) are then programmed with a series of eight SET rampdown pulses and measured. The initial pulse sequence contains eight 50-ns steps with decreasing amplitude. If the conductances obtained after a write operation result in a PCM weight within  $\pm 1.2 \mu\text{S}$  of its target, then we stop programming that device (the verify stage of the closed-loop tuning); otherwise, we reset that PCM device again (three high-amplitude 500-ns pulses) and try another pulse length. At each stage, many PCM devices are removed from consideration (for the remainder of that transfer operation), so that by the last iteration only a small number of devices are being programmed and measured. Eight-step SET rampdowns ranged from a step-width of 1.7 ns to 550 ns, with a median of 50 ns. Extended Data Fig. 7 shows cumulative conductance distributions after these various set pulses are applied to freshly reset devices. Rampdowns with fewer steps show similar results. At each iteration, we choose a new pulse-step duration on the basis of measured conductances and previously used pulses, tracking the shortest duration that was too long and the longest duration that was too short. This bisectional search results in a maximum of six pulse sequences on any given device, with 70% of devices finishing after three iterations. Extended Data Figs. 8 and 9 show histograms and correlation maps for various quantities before and after the transfer operation, taken from the last transfer operation during the MNIST experiment (Fig. 4).

Even though the volatile 3T1C conductances are critical to reach software-equivalent accuracies, the non-volatile PCM devices are more than sufficient for preserving the trained weights permanently. After performing a typical transfer onto PCM devices without any post-transfer tuning (and thus no contribution from 3T1C devices), test accuracy for the MNIST array after epoch 18 slips from 97.98%, but to only 97.25%. By taking more care during this final transfer, we were able to retain even higher accuracy (97.48%). Further improvements should be possible. After the CIFAR-10 experiment, accuracy actually increases from 88.04% to 88.17% when the weights are preserved solely on the PCM devices. Even higher non-volatile dynamic range should be possible with additional tiers of significance (for example, two non-volatile conductance pairs and one volatile lowest-significance conductance pair).

**Transfer learning.** In the hardware-based PCM experiments, the 2,048 neuron activations to the last fully connected layer are remapped to the 8-bit integer

neuron excitations used in our network implementation. Although activations across the 2,048 neurons ranged from 0 to 10.0, we clipped all excitations above 3.0. Software re-training confirmed that this clipping has a minimal effect on the resulting accuracies.

**SPICE modelling.** Random variation in FET threshold voltages introduces substantial asymmetry in both the up and down conductance change of 3T1C cells and their conductance–voltage characteristics and can therefore degrade neural-network accuracy. Unlike modelling of PCM and other emerging NVM devices, modelling of transistor characteristics and variability is both well-established and highly accurate. We created 1,000 different fitted models for each of these three effects from data generated through Monte Carlo circuit simulations in SPICE (Extended Data Fig. 10). The threshold voltage offset of each FET was sampled independently, with a standard deviation inversely proportional to the square-root of FET area<sup>42</sup>. These models were assigned randomly to all synapses in the array, allowing the mixed hardware–software experiment to account for the effect of this variability on neural-network training. Because each transistor within a 3T1C circuit chooses independently, there are  $10^9$  possible model combinations.

Circuit simulations of the 3T1C cell were carried out using LTSPICE (<https://www.linear.com/solutions/1066>). NMOS and PMOS FET model files for 90-nm technology were obtained from the Predictive Technology Model website (<http://ptm.asu.edu/>)<sup>43</sup>, which provides free-to-use BSIM models for multiple technology nodes. Oxide thickness was increased to 1.9 nm for lower gate leakage, but the model was otherwise unchanged.

From circuit simulations, we extracted (Extended Data Fig. 10): (i) the dependence of the conductance (sensed as a read current through the read FET at a constant read voltage) on capacitor voltage  $V_g$ ; (ii) the increase in  $V_g$  as a function of instantaneous voltage for multiple ‘up’ pulses; (iii) the decrease in  $V_g$  as a function of instantaneous voltage for multiple ‘down’ pulses; and (iv) the decay of  $V_g$  over time. The decay is mostly unaffected by variations in threshold voltage, so all 3T1C devices were modelled with the same decay constant of 5.16 ms.

**Power assessment.** The average power per example of the neural network was assessed by first estimating the total energy across all of the individual components and then dividing by the time per example. The core peripheral circuitry required for neural-network training was designed in IBM CMOS 9FLP (90-nm) technology; all relevant static and switching energies were obtained for all required circuitry from circuit-level simulations within the Cadence design environment. Although the capacitances of the short interconnect wiring between FETs in the peripheral circuits are not considered, we include parasitic and gate capacitances of the transistors themselves and all static and leakage power in the peripheral circuitry. Our implementation of the capacitive unit cell in 90-nm node uses the gate capacitance of a MOSFET. Because we need to use thick oxide gates for low leakage, the effective capacitance per unit area is lower ( $4.8 \text{ fF } \mu\text{m}^{-2}$ ), leading to a gate area of  $2.2 \mu\text{m}^2$  for the capacitor in this initial design ( $10.5 \text{ fF}$ ).

Extrapolating to a future, yet feasible, chip solution, we assume that a complete forward, reverse and weight-update operation could be completed in 240 ns. Because forward propagation must be completed in 80 ns for a three-layer network, each layer is assumed to have a maximum allowed pulse duration of 20 ns plus 6.7 ns of circuit set-up time. Read current from  $g^{\text{shared}}$  devices—in extra columns for forward propagation and in extra rows for reverse propagation—is scaled down before being shared. The relative significance of PCM versus 3T1C is implemented with different scaling factors on the peripheral circuitry, with associated energy fully represented in these estimates.

Energy is consumed in the array when driving the long word lines and bit lines to their appropriate voltage levels (intrinsic and loading capacitances on the wires), and in the synaptic unit cells (the read/write currents and the time for which it flows). For forward and reverse propagation, we know from our experiment both the average conductance for the PCM cells ( $3 \mu\text{S}$ ) and the average read current for the 3T1C cells ( $3 \mu\text{A}$ ). On the basis of data collected from our neural-network experiments, we compute pulse widths of the scaled average neuron excitation (for example,  $x$ ) of 6.0 ns and of the average neuron error (for example,  $\delta$ ) of 0.6 ns, relative to the maximum pulse duration of 20 ns. Our read energy results are cycle-accurate to the exact training runs in our experiment, with the only assumed parameter being the 20-ns absolute duration of the maximum-length pulse in an eventual hardware implementation. We also calculate the total number of weight updates, occasional resets and transfer operations across 20 epochs, with triage included, and then prorate these numbers per training example.

The effective number of transfers per example was calculated on the basis of statistics collected over 20 epochs of 60,000 examples each. Each transfer operation operates on a single column at a time and is assumed to involve up to nine read operations on all synapses in the column and up to eight write (partial-set) pulses on some of the devices in that column. Set pulses are assumed to be 4 ns in duration, with the set current 10 times the PCM read current. The energy for initializing the  $g$  and  $g^{\text{shared}}$  cells and for post-transfer tuning of  $g$  is calculated on the basis of a similar prorated approach. No analogue-to-digital converters are used

in the transfer process, or in the forward/reverse propagation. The same circuitry that generates pulse widths on the basis of neuron activation is used to generate a pulse that dictates the strength of the partial-set pulse used in transfer. The energy of these circuits is included in the calculations.

We assume that all transfer operations in the 2PCM + 3T1C design can be hidden within this processing time (such as while the network is performing operations on a different layer of the network). This would preclude the use of the relatively long rampdown pulses used here, mandating a closed-loop PCM conductance-tuning procedure that emphasizes short, low-amplitude partial-set and partial-reset pulses. Transfer would be distributed in time along different columns, rather than performed all at once over the entire array, as was necessary in our experiments. This choice relaxes the time constraint. Networks with a large number of layers would provide substantial time for such operation because the signal needs to pass through a large number of layers, so many cores would be idling. Considering our MNIST network (528–250–125–10 neurons) and a transfer interval of 8,000 images, the time available for transferring weights to PCM would be  $240 \text{ ns} \times 8,000 / (250 + 125 + 10) = 5 \mu\text{s}$ , where  $250 + 125 + 10$  represents the number of columns in the network. Although larger networks enable even longer programming times and relax the timescale for update, any scheme for pipelining computation so that all arrays are computing all the time would require some accommodation for weight transfer. The optimization of such a tuning procedure will be an important subject of future work.

We find the total dynamic energy per example to be 12.9 nJ for the 2PCM + 3T1C design. The primary contribution is from the first-stage forward propagate (6.3 nJ). This is consistent with the fact that in the 3T1C cell the read currents are high (of the order of microamps); the write currents are an order of magnitude lower, owing to the different current paths for write and read. This contribution is followed by the energy to route signals between stages (2 nJ) and the energy for transferring conductance information from the lower- to higher-significance pair (1.8 nJ). The corresponding average dynamic power is 53.75 mW. Static power is 0.12 mW and so the total power is 53.87 mW. Dynamic energy for the 2PCM design is 5.25 nJ, which corresponds to an average dynamic power of 21.9 mW. Static and leakage power is about the same as for the 2PCM + 3T1C design, yielding a net average power of 22.02 mW.

The effective number of multiply–accumulate operations performed is  $2(529 \times 250) + 3(251 \times 125 + 126 \times 10) = 362,405$ . The factor of 3 comes from the three passes through the hidden and output layers, first for forward inference, then for backpropagation and finally for weight update. The first layer has a pre-factor of only 2 because backpropagation through the first layer is not needed. This corresponds to a computational energy efficiency of  $28,065 \text{ GOP s}^{-1} \text{ W}^{-1}$ .

We use here a communication of neuron excitations between the device arrays that represents network layers that appears to be much more power and time efficient than is analogue-to-digital conversion, incurring approximately 2.2 nJ per training example (already included in the above analysis). Analogue-to-digital conversion and digital routing would instead incur 17.6 nJ per training example<sup>12</sup>, given the specified 12.5 million samples per second at 9 bits per samples incurring  $240 \mu\text{W}$  or  $19.2 \text{ fJ}$  per channel, which would substantially increase the total energy costs of training. The analogue-to-digital converter is needed  $250 + 125 + 10 + 125 + 250 = 760$  times to train each example. Similarly, digital-to-analogue or, more accurately, 9-bit-to-duration conversion incurs only  $0.369 \text{ fJ}$  per channel, but is used  $528 + 250 + 125 + 10 + 125 + 250 = 1,288$  times to train each example. This results in  $0.475 \text{ nJ}$  per example for a digital-to-analogue converter to reinsert digital data back into the crossbar arrays.

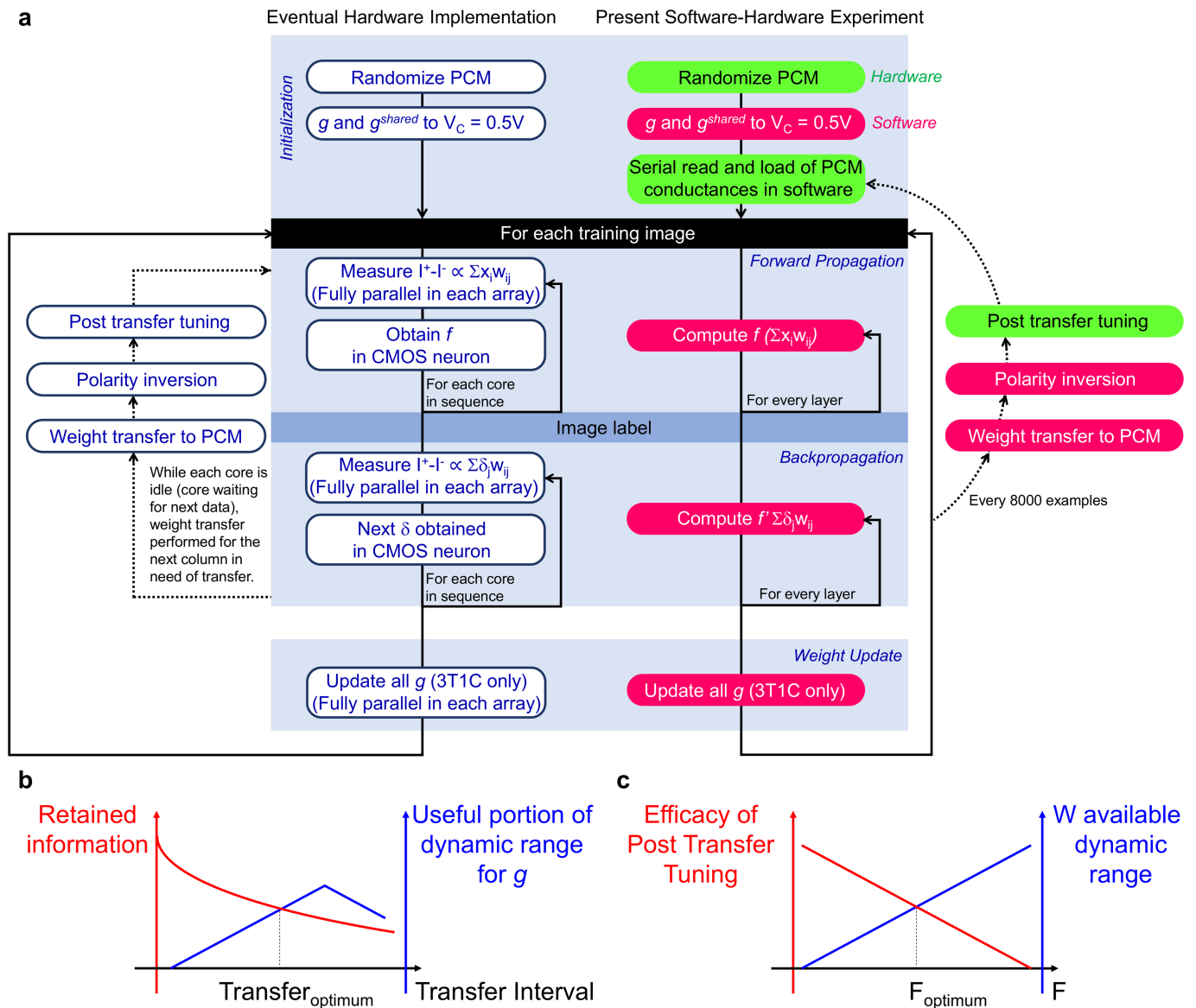
Area usage is estimated from the full layout of a  $512 \times 512$  array at 90 nm including all peripheral and routing circuits needed to perform all operations, including transfer, polarity inversion, neuromorphic read and write and array-to-array data routing. This layout incurs  $5.8 \times 10^6 \mu\text{m}^2$  at 90 nm (57% area efficiency). Assuming usage of 262,144 operations once each 240 ns (for example, 20-ns integrations occurring with an average duty cycle of about 10% for other layers) and extrapolating the area of the design to the 14-nm node through the square of the ratio of minimum wire half-pitches ( $(140 \text{ nm}/32 \text{ nm})^2$ ), we estimate a performance-per-area metric of  $3,582 \text{ GOP s}^{-1} \text{ mm}^{-2}$ . (Scaling to the 14-nm node necessitates the implementation of PCM or some other suitable NVM in that technology, posing considerable hurdles for device scaling and for delivering switching currents and voltages to the analogue memory devices.) For the V100 GPU, we assume that the best-case performance available for training is the full 30 TOP of FP16 compute. This is both pessimistic (probably only a portion of the V100 area is used in FP16 compute) and optimistic (GPU and digital accelerator performance on the fully connected layers of interest here tend to be limited by memory bandwidth to well below the peak specified performance<sup>6</sup>). In comparison, the analogue approach discussed here could potentially reduce the off-chip communication to only the incoming training examples and labels and occasional weight updates and overrides. This elimination of back-and-forth off-chip traffic for all weights, for weight-update data and for intermediate neuron excitations for all neural-network layers is the source of much

of the energy advantage over GPUs. However, our energy estimates do not include the costs of delivering input data from off-chip to the first layer (and of delivering the label data to the last layer). The importance of these costs will depend strongly on the size of the network. If a wide but shallow network with many inputs and not much computation is implemented with these techniques, then these off-chip input–output costs will probably dominate. For a large network with a reasonable number of inputs and labels, we expect that the attractive energy efficiencies described above will greatly improve the overall energy efficiency of the system.

**Data and code availability.** The training and test datasets used here are publicly available, as described in the text and the relevant references<sup>23,24</sup>. TensorFlow (<https://www.tensorflow.org>), the SPICE simulator and the PTM transistor models used here are publicly available, as per the text and relevant reference<sup>43</sup>. The specific neural-network models and parameters are provided in the text. The specific SPICE and Tensorflow decks are available on request. However, the code for our

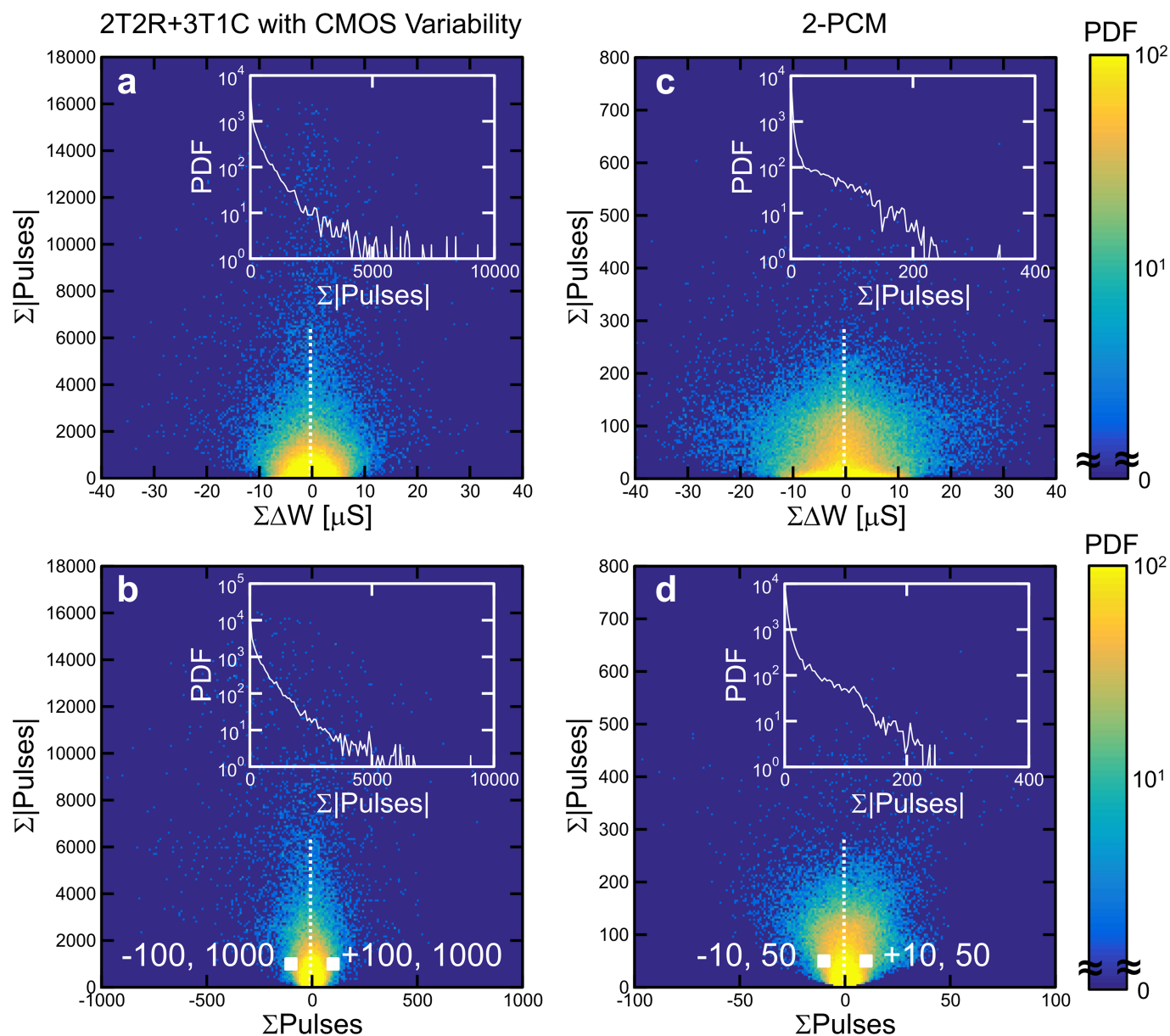
custom neural-network simulator cannot be publicly released without IBM management approval and is restricted for export by the US Export Administration Regulations under Export Control Classification Number 3A001.a.9.

40. Narayanan, P. et al. Reducing circuit design complexity for neuromorphic machine learning systems based on non-volatile memory arrays. In *2017 IEEE International Symposium on Circuits and Systems* 1–4 (IEEE, 2017).
41. Ielmini, D., Lacaíta, A. L. & Mantegazza, D. Recovery and drift dynamics of resistance and threshold voltages in phase-change memories. *IEEE Trans. Electron Dev.* **54**, 308–315 (2007).
42. Pelgrom, M. J. M., Duinmaijer, A. C. J. & Welbers, A. P. G. Matching properties of MOS transistors. *IEEE J. Solid-State Circuits* **24**, 1433–1439 (1989).
43. Cao, Y. What is predictive technology model (PTM)? *SIGDA News*. **39**, 1 (2009).
44. Bengio, Y., Louradour, J., Collobert, R. & Weston, J. Curriculum learning. In *Proc. 26th Annual International Conference on Machine Learning* 41–48 (ACM, 2009).



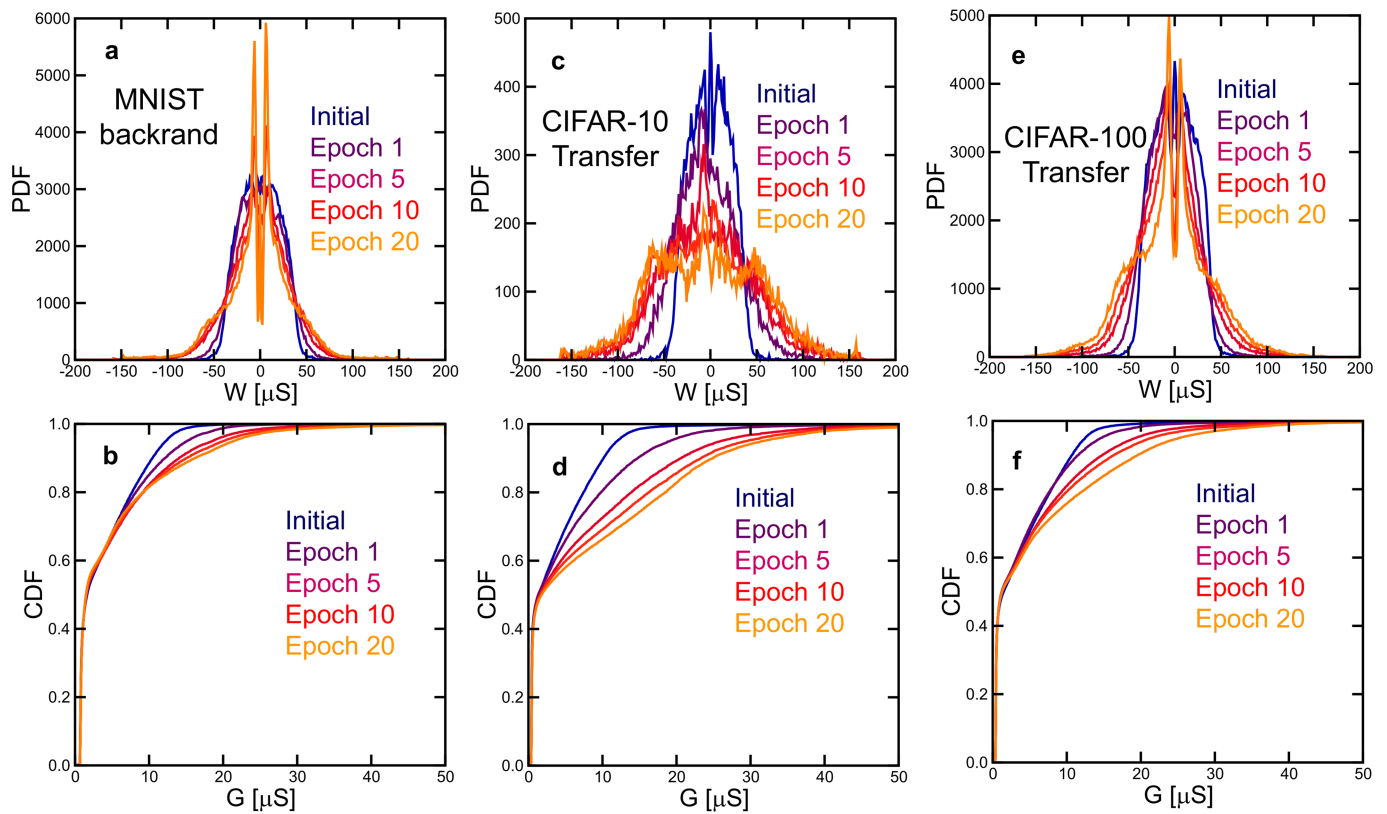
**Extended Data Fig. 1 | Flow chart comparing eventual and currently implemented DNN acceleration approaches. a,** Comparison between an eventual analogue-memory-based hardware implementation and our mixed software-hardware experiment. Although we do not implement CMOS neurons, we mimic their behaviour closely. In both schemes, weight update is performed on only the 3T1C  $g$  devices, and these contributions are later transferred to the PCM devices ( $G^+$  and  $G^-$ ). Owing to wall-clock throughput issues in our experiment, we have to perform all of the weight transfers at once. By contrast, in an eventual hardware implementation, weight transfer would take place on a distributed, column-by-column basis. Ideally, transfer for any weight column would be performed at a point in time when the neural-network computation, focused on some other layer, leaves that particular array core temporarily idle. **b,** Guidelines for optimizing the choice of transfer interval, depending on the time constant of the capacitor and the dynamic range of  $g$ . Because training of one image is performed in 240 ns, training of 8,000 images is performed in  $8,000 \times 240 \text{ ns} = 1.92 \text{ ms}$ , which is a substantial fraction of the time-constant of the capacitor (5.16 ms). Despite allowing more of the dynamic range of  $g$  to be used, a longer transfer interval would probably suffer from

poor retention of information in any volatile  $g$  device. However, even in the ideal case of an infinitely-long time constant, the transfer interval would still need to be limited, owing to the finite dynamic range of  $g$ . A long transfer interval would probably result in  $g$  values saturating owing to weight updates, leading to loss of training information before transfer. **c,** Guidelines for optimizing the choice of gain factor  $F$ . We define 'efficacy of post-transfer tuning' as the inverse of the overall residual error after  $g$  tuning. Because a larger gain factor  $F$  means more available dynamic range for each weight, larger  $F$  is desirable. However, large  $F$  also amplifies any programming errors on the PCM devices due to intrinsic device variability and limits the correction that  $g$  can provide during post-transfer tuning. The efficacy would definitely decrease monotonically, although perhaps not linearly as is sketched here. The value we chose ( $F = 3$ ) represents a reasonable trade-off for the PCM and 3T1C devices used here. For other situations,  $F$  can be initially estimated as  $F = DR_g/\sigma$ , where  $DR_g$  is the  $g$  dynamic range and  $\sigma$  is the standard deviation of the PCM programming error. Additional optimization comes with neural-network training, which includes the weak effect of drift contribution.



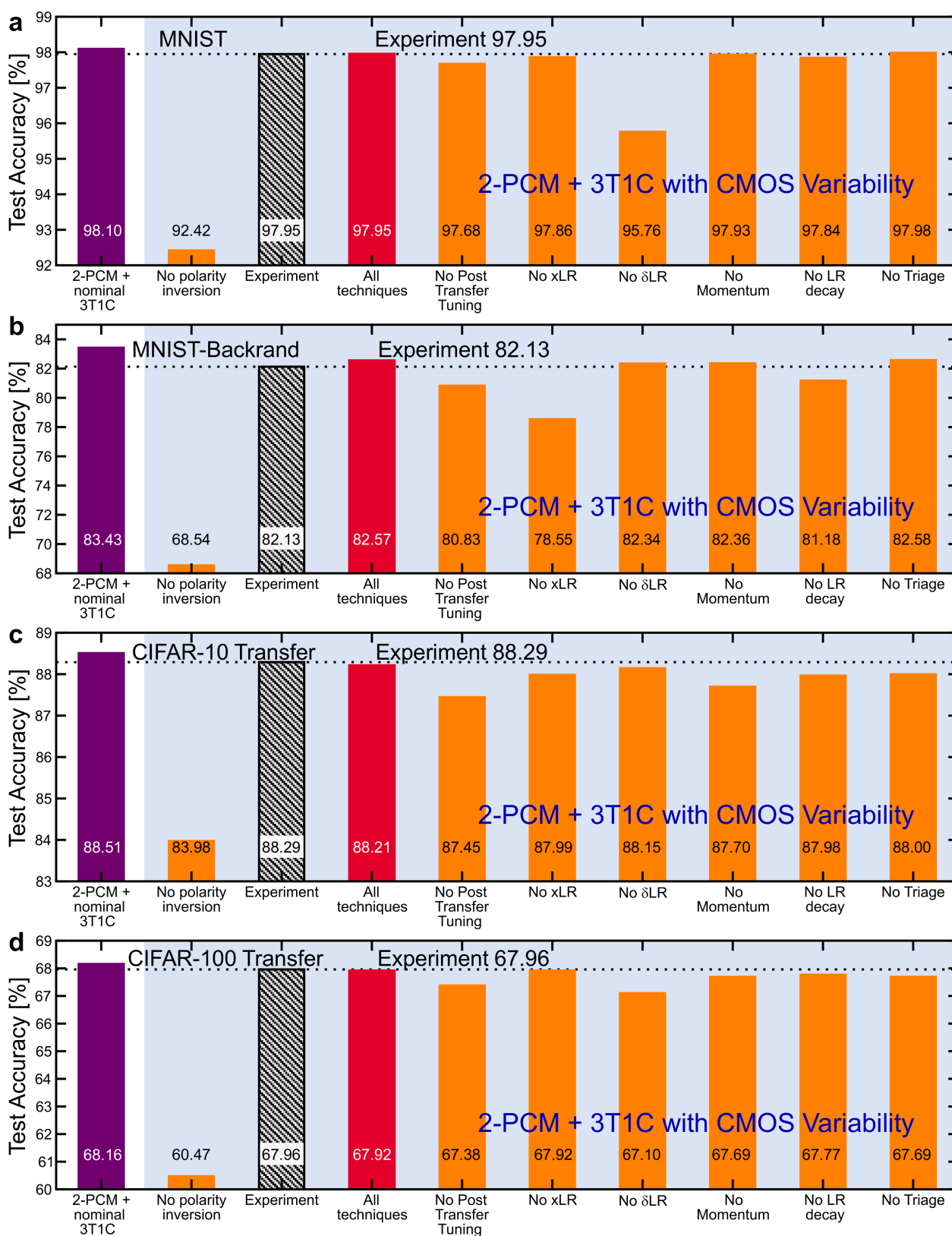
**Extended Data Fig. 2 | Weight-update requests and resulting net weight change observed during neural network training. a–d,** Simulation results based on MNIST 20-epoch simulations for the 2PCM + 3T1C cell with full CMOS variability and transfer polarity inversion (matched with the experimental results; **a, b**) and for the 2PCM cell (**c, d**). **a, c,** Correlation between the aggregate weight update across 16,000 training images (for 2PCM + 3T1C, this corresponds to two consecutive

transfer intervals) and the total number of pulses applied to obtain this weight update. **b, d,** Correlation between the aggregate number of pulses and the total number of programming pulses applied. The points chosen for Fig. 3 ( $\pm 100, 1,000$  for 2PCM + 3T1C and  $\pm 10, 50$  for 2PCM) represent typical values requested by the backpropagation algorithm. Insets show vertical cross-sections at  $\Sigma \Delta W = 0$ , where the aggregate sum of all individual weight changes  $\Delta W$  is zero (sum of pulses is zero).



**Extended Data Fig. 3 | Experimental distributions for different datasets.** (Extension of Fig. 5.) **a–f**, Weight probability density functions (PDFs) and cumulative distribution functions (CDFs) of device conductances for MNIST-backrand (**a**, **b**), CIFAR-10 transfer learning

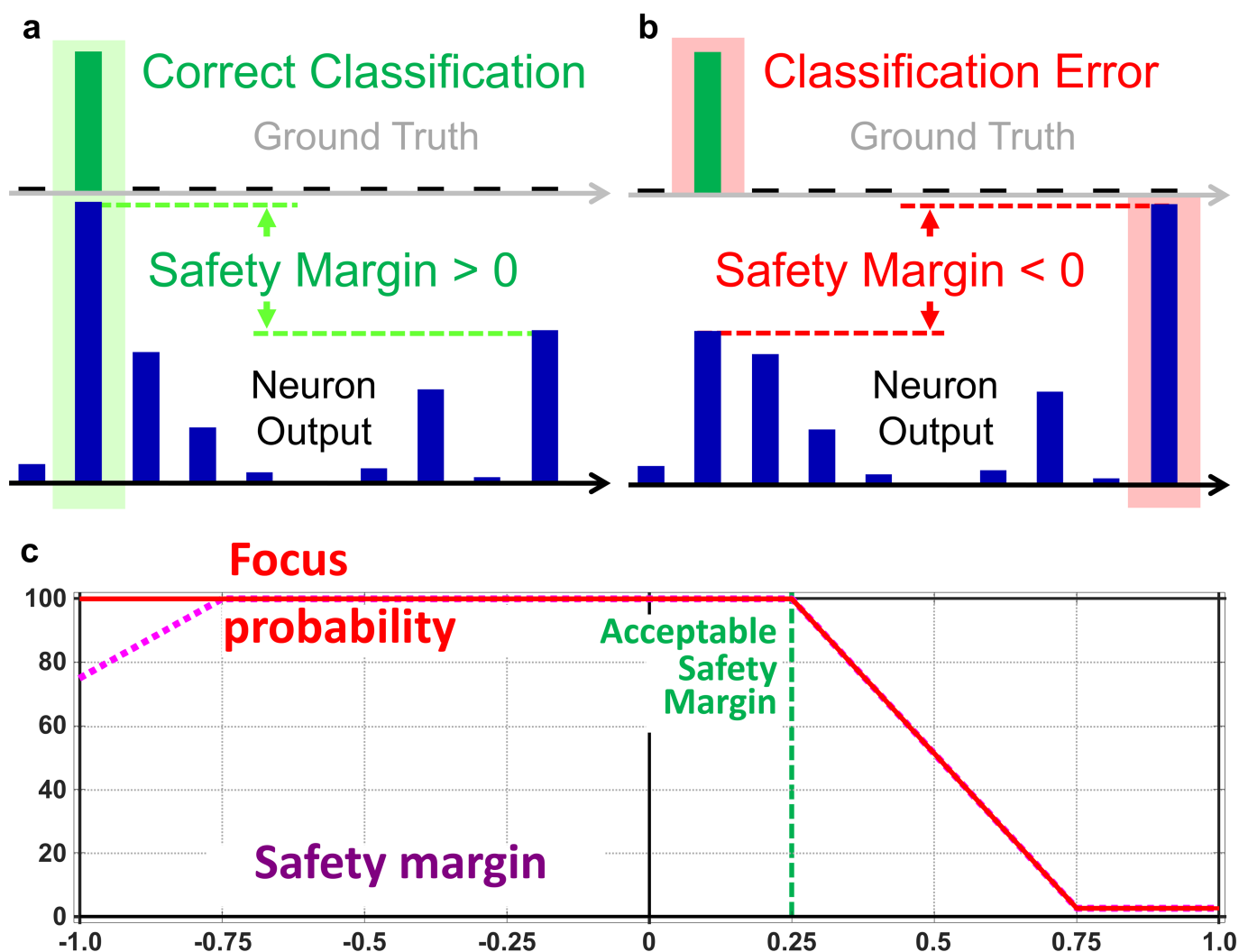
(**c**, **d**) and CIFAR-100 transfer learning (**e**, **f**). Results are shown for the initial condition and increasing epochs, from 1 to 20. For the CIFAR-100 experiment only, we increased the transfer interval to 16,000 images to reduce the overall wall-clock time.



Extended Data Fig. 4 | See next page for caption.

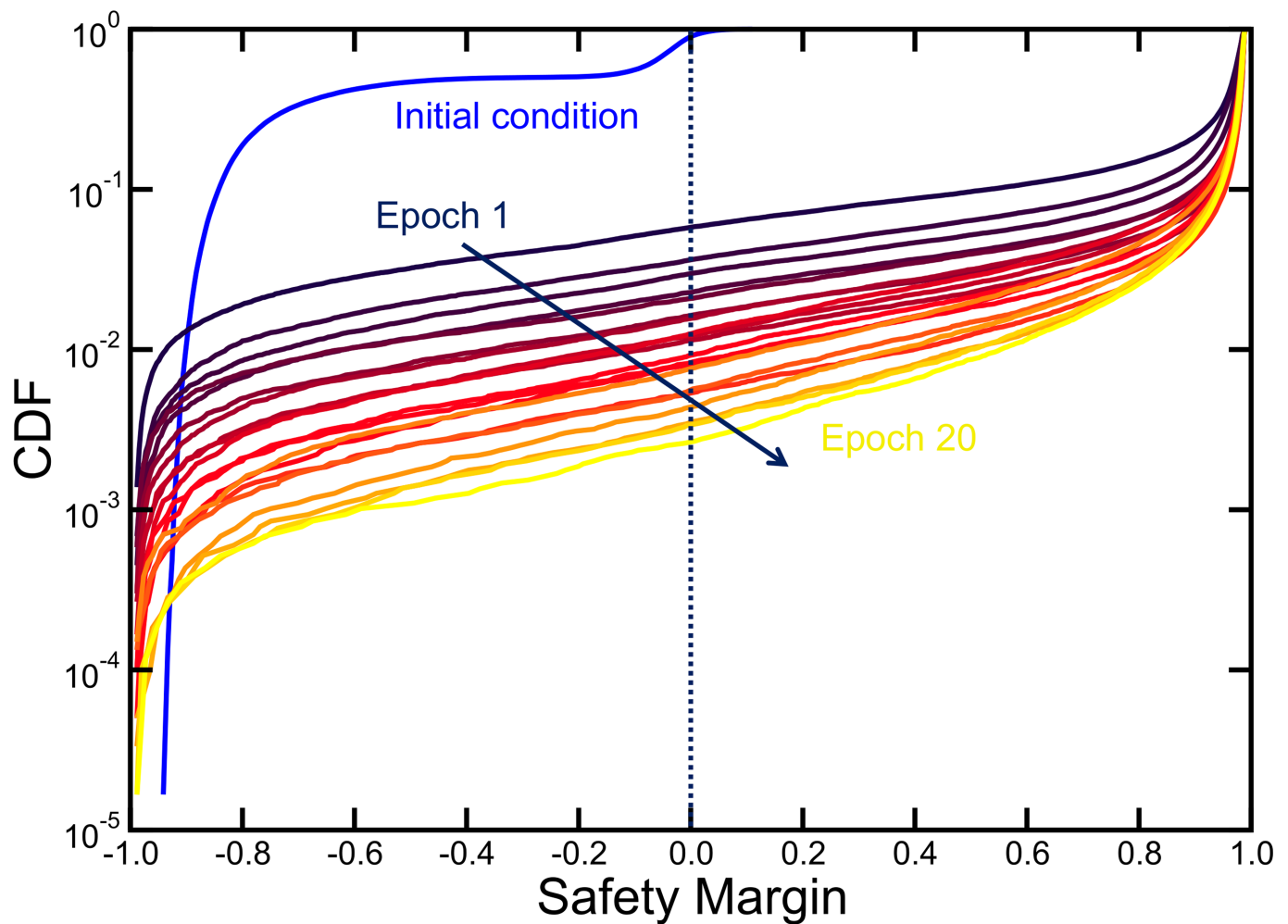
**Extended Data Fig. 4 | Effect of different techniques on neural-network training.** (Extension of Fig. 6.) **a–d**, Simulation results as in Fig. 6b, extended to all experiments performed: MNIST results (as in Fig. 6b; **a**), MNIST-backrand (**b**), CIFAR-10 transfer (**c**) and CIFAR-100 transfer (**d**). We introduce two parameters,  $x_{LR}$  and  $\delta_{LR}$ , to modify the crossbar-compatible weight-update scheme from its original conception<sup>10</sup>. The upstream neurons fire a number of weight-update pulses based on the  $x$  input signal, the global learning rate  $\eta$  and the  $x_{LR}$  coefficient; downstream neurons fire pulses depending on the error signal, the global  $\eta$  and new  $\delta_{LR}$  coefficient.  $x_{LR}$  and  $\delta_{LR}$  are both constant throughout training:  $x_{LR}$  enables

differentiation between upstream and downstream pulsing, but is constant across all layers;  $\delta_{LR}$  enables careful tuning of the importance of  $\delta$  for each weight layer.  $x_{LR}$  modulation can provide substantial accuracy benefits for MNIST-backrand (**b**) and  $\delta_{LR}$  modulation is beneficial for CIFAR-100 and particularly for MNIST (**a**, **d**). Although momentum and learning-rate (LR) decay are commonly used techniques<sup>33</sup>, their absence would not have greatly affected our experimental results. Example triage mostly provides a wall-clock advantage, but also a slight improvement in accuracy for CIFAR-10/100 transfer learning by avoiding ‘useless’ weight updates.



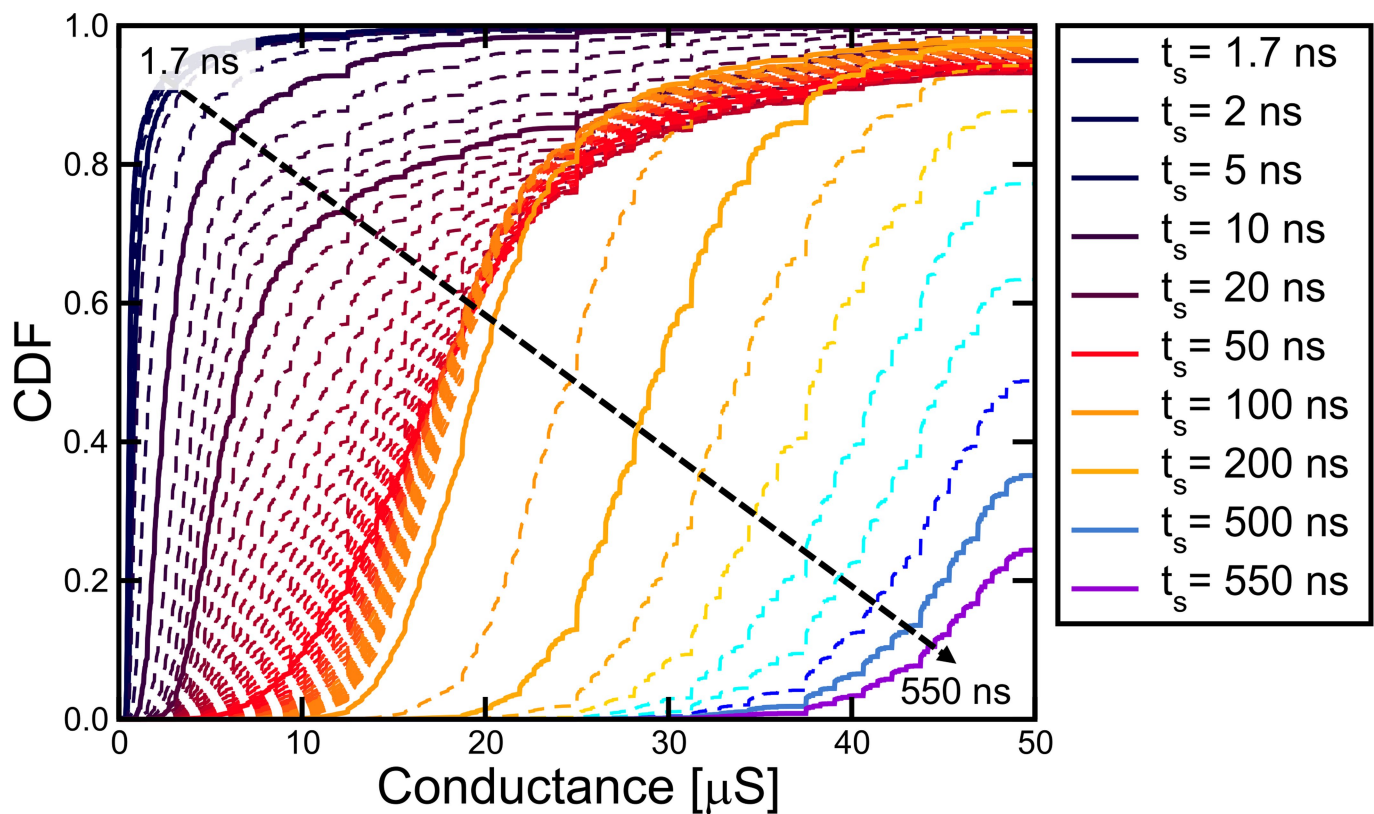
**Extended Data Fig. 5 | The safety-margin concept.** **a**, When the network classifies the output correctly (for example, the highest neuron output matches the highest ground truth), the safety margin is the positive difference between the correct neuron and the next-largest neuron. **b**, When the classification is incorrect, the safety margin is a negative number that indicates the gap by which the output neuron failed to be the highest neuron value. Preferably, we would like to calculate the safety margin for every image in each epoch, because safety margins change after each backpropagation. This is the choice made within our experiment; in a full-chip implementation of analogue-memory-based neural-network hardware accelerator with an effective minibatch size of 1, this would be fairly straightforward. Alternatively, either for minibatch-based training or for analogue hardware, we envision using a highly pipelined copy of the network designed for fast forward inference to compute safety

margins using a recent copy of the network weights. These slightly 'stale' safety margins could then be used to implement example triage. **c**, Focus probability from 0% to 100% as a function of safety margin defined from  $-1$  to  $1$ . For all safety margins below some 'acceptable' threshold, the probability of choosing to perform backpropagation on this training example is 100%. As the safety margin increases above the acceptable threshold, the focus probability decreases linearly to a non-zero minimum focus probability, to ensure that some number of already well-learned images are also backpropagated despite their high safety margin. The mapping of safety margin to focus probability can be changed during training. In addition, reducing either the focus probability or the learning rate for examples with large negative safety margins (pink dotted line) avoids damage to overall generalization in pursuit of training examples that the network may never be able to successfully classify.



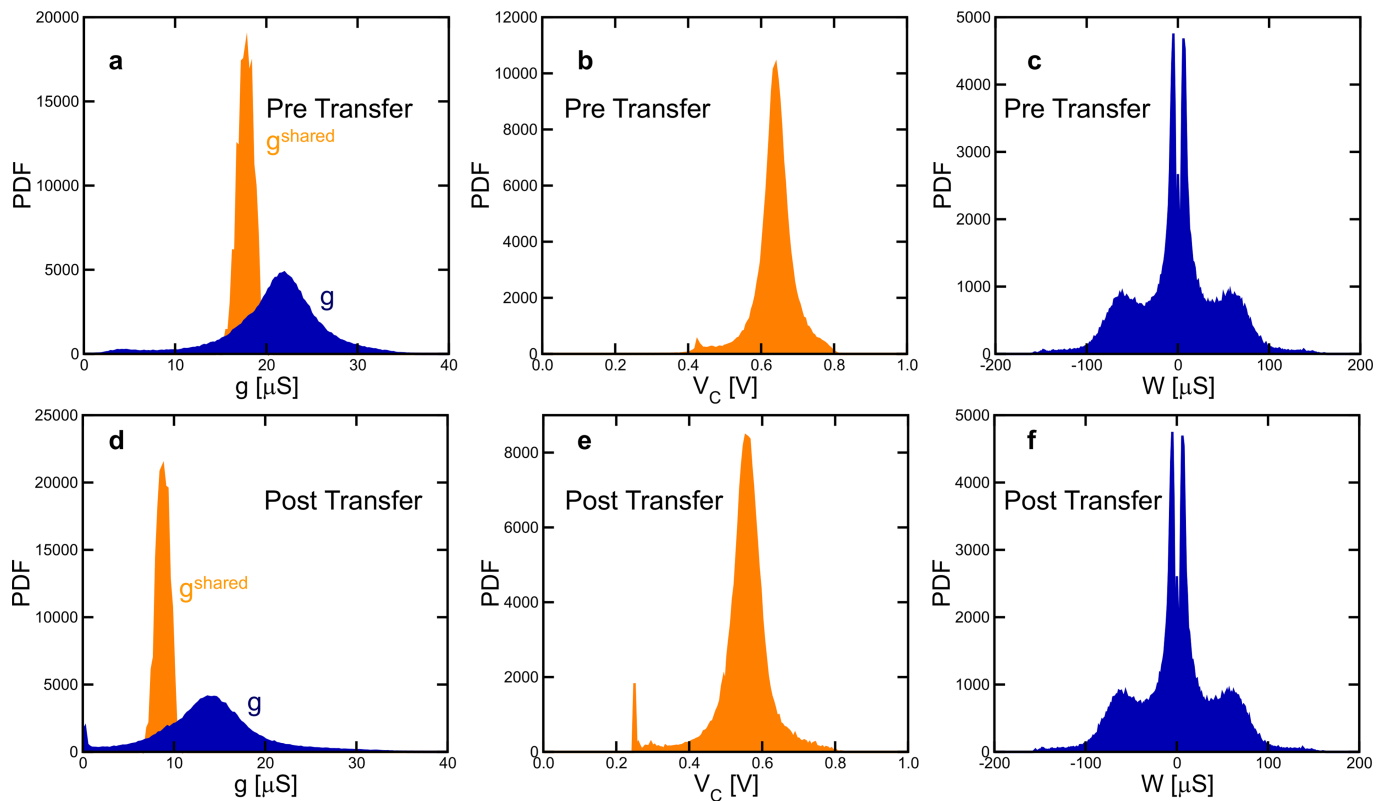
**Extended Data Fig. 6 | Safety-margin evolution during training.** During training (shown here for MNIST), the cumulative distribution of the safety margin shifts to the right, as training improves performance on the training examples. The intercept at a safety margin of zero represents the training error. Example triage can be thought of as the realization that the network does not need to train on all of the examples in the far right of this cumulative distribution, but should instead focus on those at small positive safety margins and below, with only a few training examples chosen from among those at high safety margins. The farther the safety margin distribution moves to the right, the more of an acceleration factor

that example triage can provide. Example triage can be considered a form of curriculum learning<sup>44</sup> based on the safety margin, as a highly accurate analogue measure of the current degree of certainty of the neural network. However, a substantial difference is that curriculum learning focuses on the beginning of training, with the philosophy of starting with easy examples and moving to difficult training examples. By contrast, example triage becomes effective only once the network shows some degree of performance on the training set, and is then designed to skip over easy examples in favour of difficult training examples.



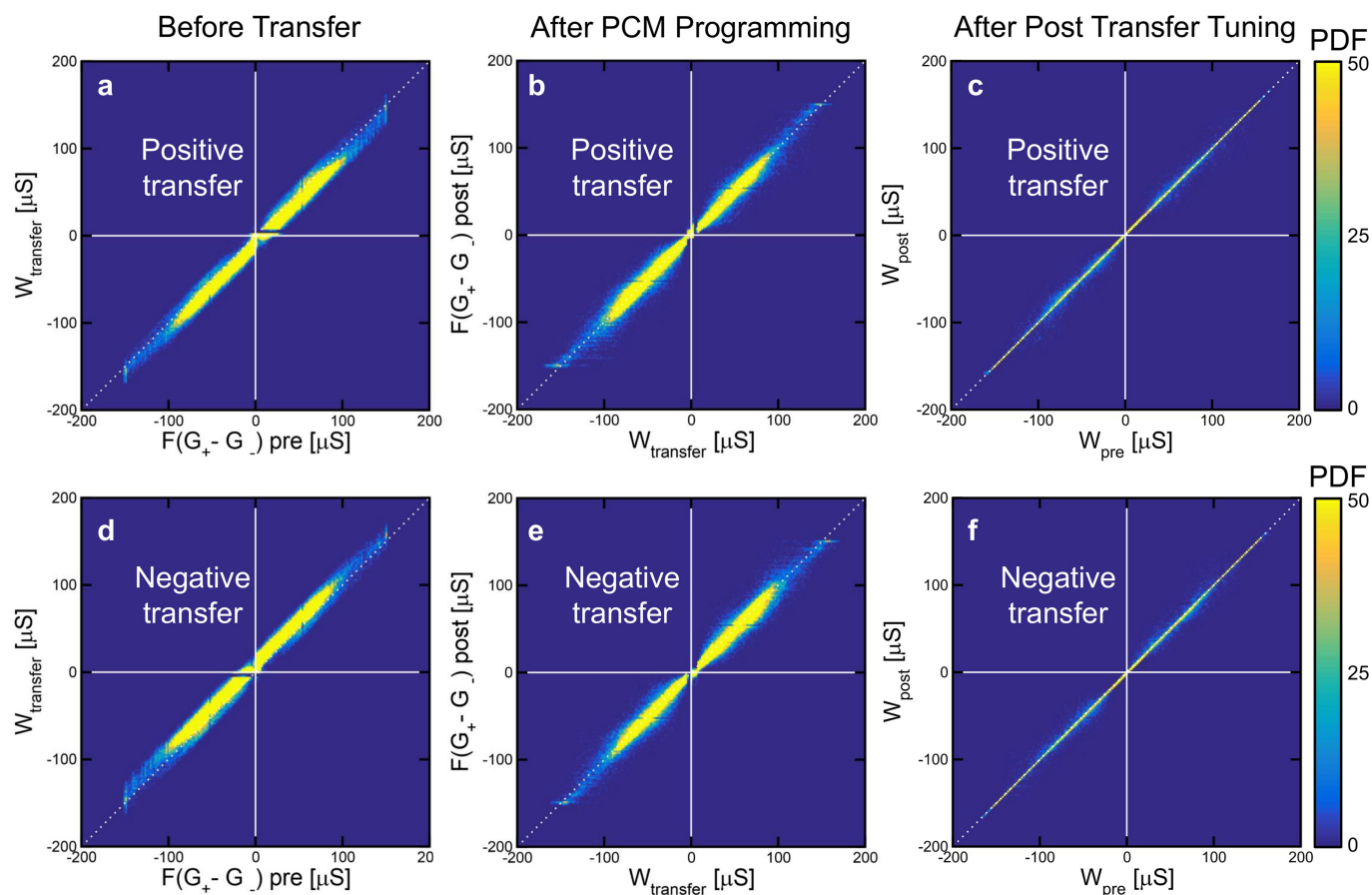
**Extended Data Fig. 7 | Experimental PCM programming distributions.** The measured cumulative distribution function of the conductances of  $512 \times 1,024$  devices programmed from full reset state with eight-step set transition rampdown pulse sequences ranging from 1.7 ns to 550 ns in step-size (for example, from 13.6 ns to 4.4  $\mu\text{s}$  in total duration) is shown.

Even though the degree of control is worse for high conductances (above 20  $\mu\text{S}$ ), to the extent that the monotonicity of the mapping from duration to conductance is disrupted, the vast majority of conductances are programmed to conductances below 20  $\mu\text{S}$  (see Fig. 4 and Extended Data Fig. 9).



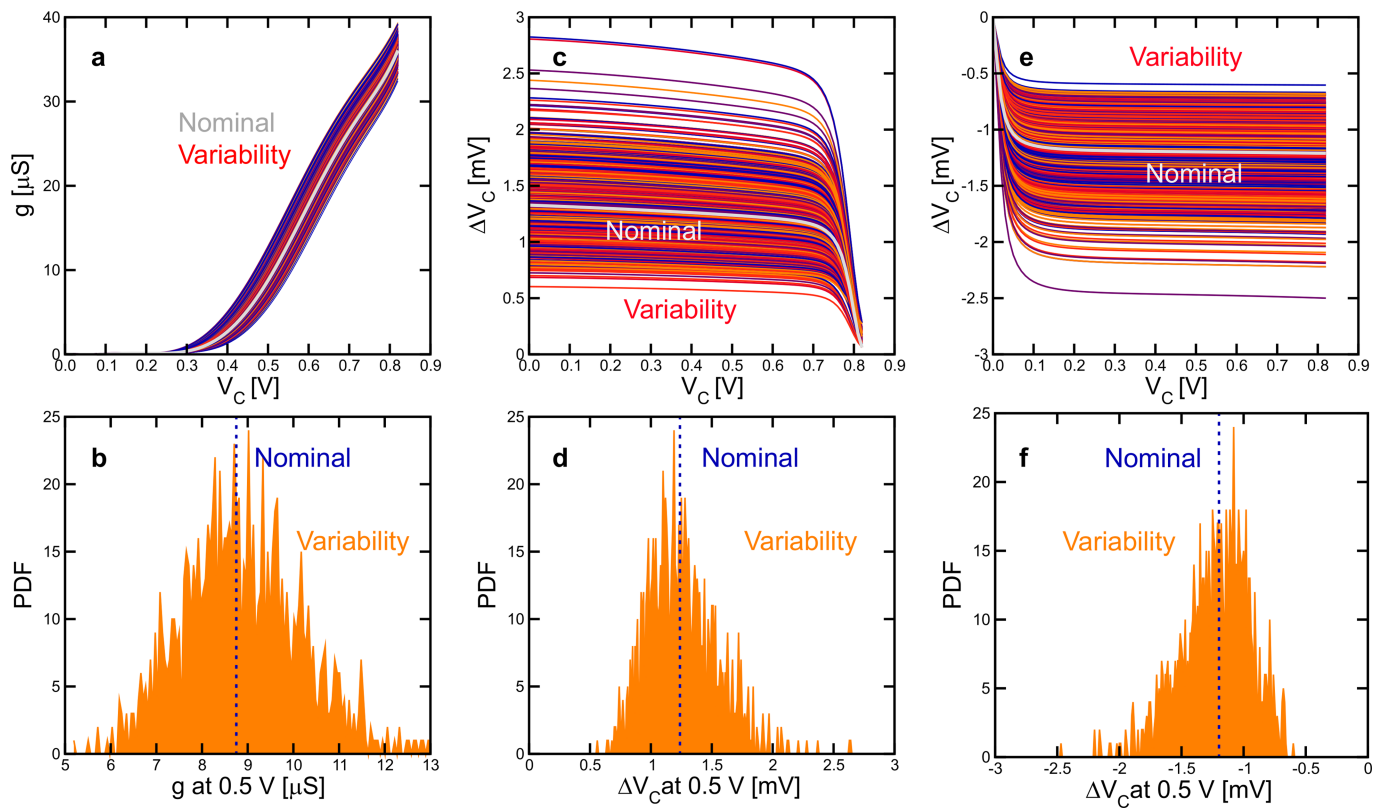
**Extended Data Fig. 8 | Analysis of weight transfer from lower- to higher-significance conductance pairs.** **a–c,** Distributions obtained before and after the last transfer in the MNIST experiment:  $g$  and  $g^{\text{shared}}$  distributions before transfer (**a**), the voltage on the capacitor of  $g$  (**b**) and the distribution of weights (**c**).  $g^{\text{shared}}$  devices are implemented as an average of the read current from three 3T1C devices for every 128 dedicated  $g$  devices to help to reduce variability. Just before transfer, the voltages on both  $g$  and  $g^{\text{shared}}$  are programmed to 0.5 V after their contribution to the weight has been extracted. **d–f,** Just after the PCM transfer, the polarity of  $g$  is inverted; the dedicated  $g$  devices are then tuned to correct the transfer error during PCM programming operation. This leads to a broad distribution of voltages on these capacitors, centred at lower voltages than just before transfer (**e**). During the long transfer interval, charge leakage in all capacitors (through both NFETs and the PFET) causes voltages to increase towards about 0.8 V. During

post-transfer tuning, the lowest voltage available to the charge subtraction circuitry is increased so that no 3T1C device can be programmed below 0.25 V (cut-off visible in **e**). Because all 3T1C conductances below that capacitor voltage are effectively zero (see Extended Data Fig. 10a), if any device were allowed to return to the weight-update operations with such an extremely low capacitor voltage, the network would be forced to fire many positive weight updates before it could effectively change that weight. Although  $g$  and  $g^{\text{shared}}$  show different shapes, the weight distribution is nearly the same as before transfer. The last transfer is shown not because it is the easiest but because it is the most important. The network has very little ability to recover from mistakes made during these last few transfers. However, data extracted for any of the other transfers throughout training would be almost indistinguishable from those shown here for the last transfer operation.



**Extended Data Fig. 9 | Effect of PCM imperfections on weight transfer.** Correlation maps obtained from the last two transfers in the MNIST experiment illustrate a typical transfer operation. The target weight  $W_{\text{transfer}}$  that we attempt to write into the PCM devices is not exactly the overall weight  $W$ , but instead  $W_{\text{transfer}} = W - \text{offset} - [g(V = 0.5 \text{ V}) - g^{\text{shared}}(V = 0.5 \text{ V})]$ . The final two terms are the residual difference between the conductances of the  $g$  and  $g^{\text{shared}}$  devices even when initialized to the same voltage, which allows the PCM devices to compensate partially for CMOS variability during transfer. The offset, equal to  $2 \mu\text{S}$ , is added because  $g$  devices are not equally good at compensating positive and negative conductance errors. At the initialization voltage of  $0.5 \text{ V}$ , device conductance is relatively small (see Extended Data Fig. 10a), providing less dynamic range to move to smaller conductances and to correct PCM devices programmed to weights that are too positive. The initial  $0.5 \text{ V}$  was chosen carefully, to accommodate substantial ‘decay’ towards  $0.8 \text{ V}$ , providing much more dynamic range for increasing 3T1C conductance. A positive offset value strongly favours negative errors, allowing us to exploit the capability for  $g$  values to increase. When  $W_{\text{transfer}}$  is positive but smaller than the offset we reset both PCM devices and use  $g$  to correct the residual error. **a**, Correlation between the weight portion encoded in PCMs before

transfer, such as  $F(G^+ - G^-)$ , with  $W_{\text{transfer}}$ . Here we expect a difference because the neural-network training has changed the weights—we now need to checkpoint these weight changes from volatile storage on the 3T1C devices into non-volatile storage on the PCM devices. **b**, Correlation between the desired  $W_{\text{transfer}}$  conductance differences and the actual  $F(G^+ - G^-)$  values obtained after PCM programming operation. With perfect devices and no offset, this should be a diagonal line along  $y = x$ . The variability we see is caused partly by PCM programming error (unintended), partly by the intentional offset and partly by CMOS initialization mismatch (where we are intentionally aiming for a ‘wrong’ PCM conductance difference to help to compensate for our flawed CMOS devices). **c**, Correlation between the weights before ( $W_{\text{pre}}$ ) and after ( $W_{\text{post}}$ ) transfer, after post-transfer tuning of  $g$  to compensate for programming errors in **b**. The goal of the transfer operation is to obtain  $W_{\text{post}} = W_{\text{pre}}$ , which would correspond to all points falling on the diagonal  $y = x$ . The effect of post-transfer tuning is clear by comparing the variability in **b** to the near-ideal behaviour in **c**. **d–f**, As in **a–c**, but for negative polarity transfer. Because the polarity of  $g$  is inverted, the offset is negative, and so the large dynamic range can be used to increase  $g$  to compensate for positive errors in PCM weight.



**Extended Data Fig. 10 | SPICE modelling of CMOS variability.**

**a–f**, Monte Carlo circuit simulations of parameter variability in 3T1C cells: measured conductance versus instantaneous voltage on the capacitor  $V_C$  (**a**); PDF of the measured conductance at  $V_C = 0.5$  V (**b**); change in voltage versus the instantaneous voltage for up pulses (**c**); PDF of change in up voltage at  $V_C = 0.5$  V (**d**); change in voltage versus the instantaneous

voltage for down pulses (**e**); and PDF of change in down voltage at  $V_C = 0.5$  V (**f**). Each graph shows data from 1,000 trials. Bold lines in **a**, **c** and **e** and dotted lines in **b**, **d** and **f** show the nominal transistor response. **a**, **b**, Variability in the read transistor whose gate is tied to the capacitor; **c–f**, variability due to variation in threshold voltage in the PMOS pull-up/NMOS pull-down FETs.

# Euryhaline ecology of early tetrapods revealed by stable isotopes

Jean Goedert<sup>1,2\*</sup>, Christophe Lécuyer<sup>1,3\*</sup>, Romain Amiot<sup>1</sup>, Florent Arnaud-Godet<sup>1</sup>, Xu Wang<sup>4</sup>, Linlin Cui<sup>4</sup>, Gilles Cuny<sup>1</sup>, Guillaume Douay<sup>5</sup>, François Fourel<sup>6</sup>, Gérard Panczer<sup>7</sup>, Laurent Simon<sup>6</sup>, J.-Sébastien Steyer<sup>8</sup> & Min Zhu<sup>9</sup>

**The fish-to-tetrapod transition—followed later by terrestrialization—represented a major step in vertebrate evolution that gave rise to a successful clade that today contains more than 30,000 tetrapod species. The early tetrapod *Ichthyostega* was discovered in 1929 in the Devonian Old Red Sandstone sediments of East Greenland (dated to approximately 365 million years ago). Since then, our understanding of the fish-to-tetrapod transition has increased considerably, owing to the discovery of additional Devonian taxa that represent early tetrapods or groups evolutionarily close to them. However, the aquatic environment of early tetrapods and the vertebrate fauna associated with them has remained elusive and highly debated. Here we use a multi-stable isotope approach ( $\delta^{13}\text{C}$ ,  $\delta^{18}\text{O}$  and  $\delta^{34}\text{S}$ ) to show that some Devonian vertebrates, including early tetrapods, were euryhaline and inhabited transitional aquatic environments subject to high-magnitude, rapid changes in salinity, such as estuaries or deltas. Euryhalinity may have predisposed the early tetrapod clade to be able to survive Late Devonian biotic crises and then successfully colonize terrestrial environments.**

Early tetrapods—four-legged vertebrates—first appeared during the Late Devonian period at around 375 million years ago<sup>1</sup>, and possibly even in the Middle Devonian (approximately 395 million years ago)<sup>2</sup>. Although they possessed limbs with digits<sup>3</sup>, detailed studies of their anatomy reveal that they inhabited aquatic environments<sup>3–5</sup>. The question of which type of aquatic ecosystem they relied upon has been closely related to the palaeoenvironmental interpretation of the Old Red Sandstone (ORS) sediments. These sediments drew the attention of palaeontologists early on, as they host a rich diversity of fossil fish<sup>6,7</sup> along with the earliest known tetrapod taxa<sup>8</sup>. Historically, the ORS sediments have been interpreted as the product of the erosion of the Caledonides mountain range<sup>9</sup> deposited in continental basins under freshwater conditions<sup>10</sup>. Fifteen years before the discovery, in the 1930s, of the first Late Devonian early tetrapods from ORS sediments, the American geologist Joseph Barrell postulated that these animals should have inhabited and evolved in the freshwater environments represented by the ORS sediments<sup>11</sup>. Thus, when the first early tetrapods *Ichthyostega* and *Acanthostega* were discovered in the ORS sediments of East Greenland, Barrell's prediction gave birth to a strong and lasting paradigm: early tetrapods and their associated fauna inhabited freshwater environments. However, this view has been challenged by many authors, who have defended the idea that ORS sediments may have been subjected to greater marine influences than previously thought, and that early tetrapods may have been able to tolerate higher salinities<sup>12–16</sup>. Since the discovery of *Ichthyostega* and *Acanthostega*, additional early tetrapod body fossils<sup>1,17–25</sup> and trackways<sup>2</sup> have been found worldwide in sedimentary deposits reflecting marine, brackish or freshwater environments<sup>26</sup>. These various interpretations have revived the debate concerning the living environments of early tetrapods and their associated fauna. However, this question remains unresolved, being hampered by the lack of direct environmental tracers that could help to decipher the living environment of early tetrapods.

Here we address this problem by applying a new isotopic tracer, the sulfur isotope composition of bone sulfates (hereafter,  $\delta^{34}\text{S}_{\text{bone}}$ ), conjointly with carbon and oxygen isotope analyses (the  $\delta^{13}\text{C}_{\text{c}}$  and  $\delta^{18}\text{O}_{\text{c}}$  of carbonates, and the  $\delta^{18}\text{O}_{\text{p}}$  of phosphates) of 51 fossilized bones from Devonian vertebrates, recovered from two geographically distant early tetrapod localities which have yielded a similar fauna (placoderms, non-tetrapod sarcopterygians and tetrapods; Extended Data Fig. 1): the Upper Devonian sequences of East Greenland and Chinese Ningxia Hui autonomous region.

## $\delta^{34}\text{S}_{\text{bone}}$ is a tracer of aquatic environments

Natural sulfur-bearing compounds have sulfur isotope compositions ( $\delta^{34}\text{S}$ ) that are highly variable among various types of terrestrial and aquatic environments<sup>27</sup>. For instance, dissolved sulfates in present-day seawater have high and relatively uniform  $\delta^{34}\text{S}$  values close to +21‰ (relative to the international reference, Vienna Canyon Diablo Troilite). Most freshwater environments (for example, rivers, lakes and ponds) have comparatively lower sulfate  $\delta^{34}\text{S}$  values, ranging from –20‰ to +20‰<sup>27</sup>. Because environmental sulfur isotope composition is recorded in animal organic tissues (for example, hairs or muscles) with little fractionation<sup>28</sup>, sulfur isotope analysis has been used in ecological studies as a tracer of living environments. It is especially suitable for indicating whether an animal is living in a freshwater or in a marine environment<sup>29</sup>. Nonetheless, sulfur is also present at low concentrations (less than 0.6 wt%, Supplementary Table 1) in the apatite crystals ( $\text{Ca}_5(\text{PO}_4)_3(\text{OH})$ ) that constitute bones—in the form of sulfate ( $\text{SO}_4^{2-}$ ) substituting for the phosphate ( $\text{PO}_4^{3-}$ ) group—and the potential for apatite to be accurately measured for its  $\delta^{34}\text{S}$  value has recently been demonstrated<sup>30</sup>.

To quantify the isotopic fractionations associated with the incorporation of environmental sulfate in apatite and organic tissues, we have analysed both organic (muscles) and mineralized (bones) tissues of

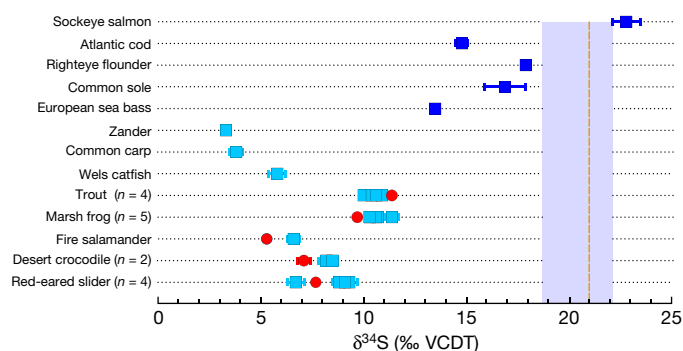
<sup>1</sup>CNRS, UMR 5276 LGL-TPE, Univ Lyon, Université Lyon 1, Ens de Lyon, Villeurbanne, France. <sup>2</sup>CNRS, UMR 5199 PACEA, Université de Bordeaux, Bordeaux, France. <sup>3</sup>Institut Universitaire de France, Paris, France. <sup>4</sup>Key Laboratory of Cenozoic Geology and Environment, Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing, China. <sup>5</sup>Zoo de Lyon, Lyon, France. <sup>6</sup>CNRS, UMR 5023 LEHNA, Université Claude Bernard Lyon 1, Université de Lyon, ENTPE, Villeurbanne, France. <sup>7</sup>CNRS, UMR 5306 ILM, Université Claude Bernard Lyon 1, Université de Lyon, Villeurbanne, France. <sup>8</sup>CNRS, UMR 7207 CR2P, MNHN-UPMC, Muséum national d'Histoire naturelle, Galerie de Paléontologie, Paris, France. <sup>9</sup>Key Laboratory of Vertebrate Evolution and Human Origins of Chinese Academy of Sciences, Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences, Beijing, China. \*e-mail: jean.goedert@ens-lyon.fr; christophe.lecuyer@univ-lyon1.fr

present-day vertebrates for their  $\delta^{34}\text{S}$  values, as well as the  $\delta^{34}\text{S}$  values of their environmental water and food (Supplementary Table 1). Vertebrates living in seawater have  $\delta^{34}\text{S}_{\text{bones}}$  (+14.8‰ to +22.8‰) and  $\delta^{34}\text{S}_{\text{muscles}}$  (+17.4‰ to +19.8‰) values higher than those of vertebrates living in freshwater environments, which have values of +3.8‰ to +11.4‰ for bones and +5.1‰ to +12.6‰ for muscles (Fig. 1 and Supplementary Table 1). These values reflect the generally high  $\delta^{34}\text{S}$  values of dissolved sulfate in seawater compared to those of sulfate in freshwater<sup>27</sup>. On the whole, the sulfur isotope fractionation between bone apatite and ambient water is of low magnitude ( $\Delta^{34}\text{S}_{\text{bone-water}} = +0.8\text{‰}$ ;  $1\sigma = 0.8\text{‰}$ ;  $n = 5$ ); this is also observed between vertebrate muscles and food ( $\Delta^{34}\text{S}_{\text{muscle-food}} = -1.8\text{‰}$ ;  $1\sigma = 0.01\text{‰}$ ;  $n = 2$ ). These values are consistent with previously reported isotope fractionations between sulfur of organic tissues and sulfur present in food sources ( $\Delta^{34}\text{S}_{\text{consumer-food}} = +0.5\text{‰}$ ;  $1\sigma = 2.4\text{‰}$ )<sup>28</sup>. These data strongly support the idea that  $\delta^{34}\text{S}_{\text{bone}}$  values directly reflect  $\delta^{34}\text{S}_{\text{water}}$  values.

We have also measured the  $\delta^{34}\text{S}$  values of two food pellets used to raise five marsh frogs (*Pelophylax ridibundus*; Supplementary Table 1) and four freshwater trout (two specimens of *Oncorhynchus mykiss* and one each of *Salvelinus fontinalis* and *Salmo trutta*; Supplementary Table 1). For these specimens,  $\delta^{34}\text{S}_{\text{bone}}$  values are closer to their corresponding  $\delta^{34}\text{S}_{\text{water}}$  values (marsh frogs:  $\Delta^{34}\text{S}_{\text{bone-water}} = +1.0\text{‰}$ ;  $1\sigma = 0.5\text{‰}$  and trout:  $\Delta^{34}\text{S}_{\text{bone-water}} = -0.9\text{‰}$ ;  $1\sigma = 0.4\text{‰}$ ) than to their corresponding  $\delta^{34}\text{S}_{\text{food}}$  values (marsh frogs  $\Delta^{34}\text{S}_{\text{bones-food}} = -2.4\text{‰}$ ;  $1\sigma = 0.5\text{‰}$  and trout  $\Delta^{34}\text{S}_{\text{bones-food}} = -2.0\text{‰}$ ;  $1\sigma = 0.4\text{‰}$ ). These observations suggest that the sulfur present in mineralized tissues in the form of sulfate derives principally from dissolved seawater sulfate. Sulfur isotope analysis of mineralized tissues thus represents a suitable method for deciphering the aquatic environment of extinct vertebrates for which only bone or tooth apatite has been preserved.

### State of preservation of Devonian apatites

During fossilization processes, secondary minerals can precipitate at the surface or within the bones (for example, pyrite ( $\text{FeS}_2$ ) and sulfate minerals ( $\text{BaSO}_4$  or  $\text{SrSO}_4$ )). These secondary minerals would tend to increase the sulfur content of bones, which is typically lower than 0.6 wt% in present-day biogenic apatites (Supplementary Table 1). The sulfur contents of the Devonian samples we studied range from 0.1 wt% to 5.8 wt% (Supplementary Table 2) with a mean value of 0.7 wt%



**Fig. 1 |  $\delta^{34}\text{S}_{\text{bone}}$  and  $\delta^{34}\text{S}_{\text{water}}$  values of modern reptiles, amphibians and fish.** Marine fish (dark blue squares) have higher  $\delta^{34}\text{S}_{\text{bone}}$  values than other vertebrates relying on freshwater (light blue squares) in their environment.  $\delta^{34}\text{S}_{\text{bone}}$  values are close to their corresponding  $\delta^{34}\text{S}_{\text{water}}$  values (red dots). The European sea bass—a marine fish that swims episodically in brackish and even fresh waters—has an intermediate  $\delta^{34}\text{S}_{\text{bone}}$  value of +13.5‰, between those of strictly marine and freshwater species. The blue shaded area surrounding the mean  $\delta^{34}\text{S}$  value of marine water corresponds to the sulfur isotope variability in marine environments<sup>27</sup>. Each dot represents a biologically independent animal ( $n = 24$ ) and corresponds to the average values of three repeated measurements. Each error bar corresponds to 1 s.d. (Supplementary Table 1). Results are given as variations in parts per mille from the ratio of  $^{34}\text{S}/^{32}\text{S}$  in Vienna Canyon Diablo Troilite (‰ VCDT). Dashed yellow vertical line, mean  $\delta^{34}\text{S}$  value of present-day seawater.

( $1\sigma = 0.8\text{‰}$ ) and a median of 0.4 wt%. When the sulfur concentration increases to values higher than 0.6 wt%,  $\delta^{34}\text{S}$  values tend to reach values around 23–24‰ (Extended Data Fig. 2a). This could reflect an incorporation of secondary sulfur-bearing minerals disturbing the pristine sulfur isotope composition. However, elemental contents (Mg, Mn, Fe, Cu, Ba and Sr, Supplementary Table 3) of Devonian apatite samples are too low to account for the presence of stoichiometric secondary mineral phases such as pyrite or sulfate minerals. Only barium and, to a lesser extent, strontium display significant correlations with sulfur (Extended Data Fig. 3). Nonetheless, the slopes of these correlation lines are too low to account for the stoichiometric slope defined by the  $\text{Ba}_x\text{Sr}_{(1-x)}\text{SO}_4$  phase (Extended Data Fig. 3e, f). This means that only a small fraction of the total sulfur present in Devonian apatites may be associated with a secondary precipitated mineral phase such as  $\text{BaSO}_4$  or  $\text{SrSO}_4$ . We also measured the elemental content of seven apatite-associated sedimentary matrices (Supplementary Table 3). They all have sulfur concentrations ten times lower than their associated fossilized apatite samples (S wt% = 0.09%;  $1\sigma = 0.04$ ; Extended Data Fig. 3). This observation argues against a possible sulfur contamination from the host sedimentary matrix.

Fourier-transform infrared spectroscopy analysis unambiguously identifies the presence of structurally substituted  $\text{SO}_4^{2-}$  in Devonian apatites (Extended Data Fig. 4). Furthermore,  $\text{SO}_4^{2-}/\text{PO}_4^{3-}$  ratios display a significant correlation with the S/P ratios calculated from elemental analyses (Extended Data Fig. 5). The intercept of the calculated regression is lower than 0 (that is,  $\text{SO}_4^{2-}/\text{PO}_4^{3-} > \text{S/P}$ ). This is probably explained by the fact that some residues of apatite powders may not be entirely dissolved before elemental analysis. By contrast, Fourier-transform infrared spectroscopy analysis does not require any chemical pretreatment of apatite powders that may involve a loss of matter before analysis. The slope of the correlation for all the samples is equal to 0.60 ( $R^2 = 0.50$ ,  $P = 2.6 \times 10^{-7}$ ) and it is not statistically different from 1 after removing 5 outliers ( $R^2 = 0.74$ ,  $P = 2.7 \times 10^{-11}$ ). This observation demonstrates that the elemental sulfur in samples of Devonian apatite is mainly present as sulfate that is structurally substituted for the phosphate in the apatite lattice. This observation also confirms that only a small fraction of sulfur may be associated with a secondary precipitated mineral phase such as  $\text{BaSO}_4$  or  $\text{SrSO}_4$ . Even the sample with the highest sulfur content (5.8 wt%; Supplementary Table 2) shows a very pronounced infrared peak, which undoubtedly indicates the presence of sulfate that has been structurally substituted for phosphate in the apatite lattice (Extended Data Fig. 4).

The bond dissociation energy of sulfate ( $\text{SO}_4^{2-}$ ) is elevated ( $\Delta_f H^\circ = -909.27$  kJ per mol at 298.15 K<sup>31</sup>). Therefore, structurally substituted sulfate is relatively robust regarding potential sulfur substitution. Nonetheless, we cannot disregard the possibility that samples with a sulfur content higher than 0.6 wt% (the average of present-day material) may have incorporated secondary sulfates by substitution during recrystallization processes. Indeed, the crystallinity index (Supplementary Table 5) indicated that Devonian apatite samples have undergone substantial recrystallization. For instance, this process of recrystallization may explain the significant negative correlation observed between calcium and barium (Extended Data Fig. 6). This correlation suggests that some calcium was substituted with barium during this recrystallization process. It is worth noting that the samples defining this correlation all have high sulfur content. Nonetheless, depending on the fluids interacting during the recrystallization process, chemical changes will not necessarily affect the different chemical elements of the apatite lattice in the same manner. This probably explains why there are no significant correlations between the crystallinity index and both sulfur content and isotopic ratios (Extended Data Fig. 7). The recrystallization of Devonian apatites did not systematically result in sulfate being added to the apatite by a mechanism of substitution accompanied by an alteration of the pristine sulfur isotope composition. We therefore chose to remove 19 biogenic apatite samples that had sulfur concentrations higher than 0.6 wt% (Extended Data Fig. 2b

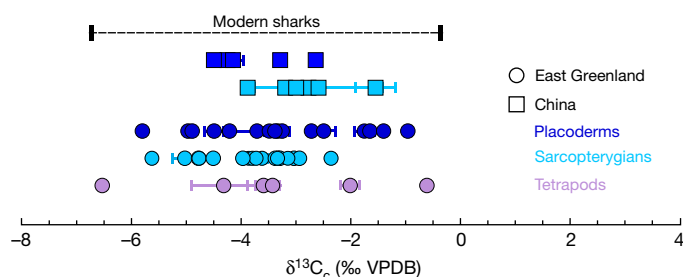
and Supplementary Table 2), as we considered them to be 'dubious' and discarded them from further interpretations.

The preservation of pristine oxygen isotope compositions of phosphate apatites was assessed using the crystallinity index and the ratio of  $\text{CO}_3^{2-}$  to  $\text{PO}_4^{3-}$  (Supplementary Table 5). The crystallinity index values of modern material range between 2.8 and 3.3<sup>32,33</sup>. In our material, the crystallinity index values range from 1.9 to 10.3, with a mean value of 5.0 ( $1\sigma = 1.4$ ). Therefore, the majority of the samples experienced substantial recrystallization during diagenesis. Nonetheless, the absence of any significant correlation between crystallinity index values and  $\delta^{18}\text{O}_\text{p}$  values (Extended Data Fig. 8;  $R^2 = 0.054$ ,  $P = 0.12$ ) proves that these processes of recrystallization did not alter the oxygen isotope composition of phosphate, as previously observed by several authors<sup>32,33</sup>. On the whole, the calcium content and phosphorus content of Devonian samples are significantly correlated ( $R^2 = 0.79$ ,  $P = 1.5 \times 10^{-14}$ ), with a slope of 1.19; this slope is close to the Ca/P ratio of the standard reference materials NIST1400 (bone ash) and NIST1486 (bone meal) (Ca/P = 1.49), and to that of present-day hydroxylapatite (Ca/P = 1.67) (Extended Data Fig. 9). These results confirm that the loss of phosphate during recrystallization was almost stoichiometric for all samples. Consequently, the loss of phosphate does not imply oxygen isotopic fractionation. This is confirmed by the  $\text{CO}_3^{2-}/\text{PO}_4^{3-}$  ratios, which range from 0.1 to 2.3 with a mean value of 0.4 ( $1\sigma = 0.4$ ). Thirty-seven samples have  $\text{CO}_3^{2-}/\text{PO}_4^{3-}$  ratios in the range of 0.15–0.7<sup>32</sup>, which indicates that there have been no important changes in the mineral fraction of apatite samples. Again,  $\text{CO}_3^{2-}/\text{PO}_4^{3-}$  values do not correlate significantly with the  $\delta^{18}\text{O}_\text{p}$  values (Extended Data Fig. 8;  $R^2 = 0.0016$ ,  $P = 0.79$ ). These two observations strongly argue in favour of at least a partial preservation of the pristine oxygen isotope compositions. Nonetheless, we also compared the oxygen isotope compositions of both bone phosphate ( $\delta^{18}\text{O}_\text{p}$ ) and bone carbonate ( $\delta^{18}\text{O}_\text{c}$ ) with the carbonate content ( $\text{CO}_3^{2-}$  wt%) of each bone sample (Extended Data Fig. 2b and Supplementary Table 2). In present-day mammals,  $\delta^{18}\text{O}_\text{p}$  values are between 7 and 9‰ lower than  $\delta^{18}\text{O}_\text{c}$  values<sup>34</sup>, although they can be as much as 14.7‰ lower in sharks<sup>35</sup>. Microbially induced diagenetic alteration has previously been shown to increase the offset between  $\delta^{18}\text{O}_\text{p}$  and  $\delta^{18}\text{O}_\text{c}$  values<sup>36</sup>. In addition,  $\text{CO}_3^{2-}$  wt% values of teeth and bones of present-day vertebrates range from 2.0 to 13.4 wt%<sup>35,37,38</sup>. We identified and discarded for further interpretation 1 biogenic apatite sample with a  $\delta^{18}\text{O}_\text{c}$ – $\delta^{18}\text{O}_\text{p}$  difference that exceeded 14.7‰ (Extended Data Fig. 2b and Supplementary Table 2), and 5 samples with  $\text{CO}_3^{2-}$  wt% values that were higher than 13.4% (Extended Data Fig. 2b and Supplementary Table 2).

### $\delta^{13}\text{C}_\text{c}$ and $\delta^{34}\text{S}_\text{bone}$ show a seawater influence

Carbon isotope compositions of apatite carbonate ( $\delta^{13}\text{C}_\text{c}$ ) from air-breathing vertebrates have previously been shown to primarily reflect animal diets, with a magnitude of  $^{13}\text{C}$ -enrichment relative to  $^{12}\text{C}$  that varies among species<sup>39</sup>. The  $\delta^{13}\text{C}_\text{c}$  values of the Devonian samples from East Greenland and China range from  $-6.5\text{‰}$  to  $-0.6\text{‰}$  (mean  $\delta^{13}\text{C}_\text{c} = -3.5\text{‰}$ ;  $1\sigma = 1.3\text{‰}$ ) and from  $-4.5\text{‰}$  to  $-1.5\text{‰}$  (mean  $\delta^{13}\text{C}_\text{c} = -3.3\text{‰}$ ;  $1\sigma = 0.9\text{‰}$ ), respectively (Fig. 2 and Supplementary Table 2). These values are comparable to those obtained for extant sharks that prey on marine mammals and fish<sup>35</sup>, thus indicating that the  $\delta^{13}\text{C}_\text{c}$  values recorded in Devonian apatites are compatible with marine-influenced diets.

The  $\delta^{34}\text{S}_\text{bone}$  values of Devonian vertebrates from East Greenland and China range from  $+12.5\text{‰}$  to  $+31.8\text{‰}$  (mean  $\delta^{34}\text{S}_\text{bone} = +21.1\text{‰}$ ;  $1\sigma = 5.6\text{‰}$ ) and from  $+14.3\text{‰}$  to  $+22.3\text{‰}$  (mean  $\delta^{34}\text{S}_\text{bone} = +17.6\text{‰}$ ;  $1\sigma = 3.0\text{‰}$ ), respectively (Fig. 3a and Supplementary Table 2). These mean values are elevated and close (a few per mille lower) to that proposed for the Late Devonian seawater ( $+25\text{‰}$ <sup>40</sup>), which indicates that the  $\delta^{34}\text{S}$  values of bone sulfur of Devonian early tetrapods and their associated fauna derive principally from Devonian seawater sulfate. The considerable variability of  $\delta^{34}\text{S}$  values is consistent with aquatic environments subjected to high-magnitude, rapid changes in salinity (Fig. 3b), and is also comparable to values that we observe in modern



**Fig. 2 |  $\delta^{13}\text{C}_\text{c}$  of Devonian vertebrate bone apatites.** For comparison, the range of  $\delta^{13}\text{C}_\text{c}$  values in extant sharks is reported<sup>35</sup>. Similar to those of extant sharks, the  $\delta^{13}\text{C}_\text{c}$  values of Devonian vertebrates are compatible with marine-influenced food sources. Each dot represents an independent Devonian apatite sample ( $n = 51$ ) and corresponds to the average value of two repeated measurements. Each error bar corresponds to 1 s.d. (Supplementary Table 2). Results are given as variations in parts per mille from the ratio of  $^{13}\text{C}/^{12}\text{C}$  in Vienna Pee Dee Belemnite (‰ VPDB).

fish populations living in estuaries<sup>29,41</sup>. As we analysed fossil assemblages that aggregate the remains of vertebrates that probably died at different periods, we cannot disregard the fact that some isotopic variability may be due to temporal variation.

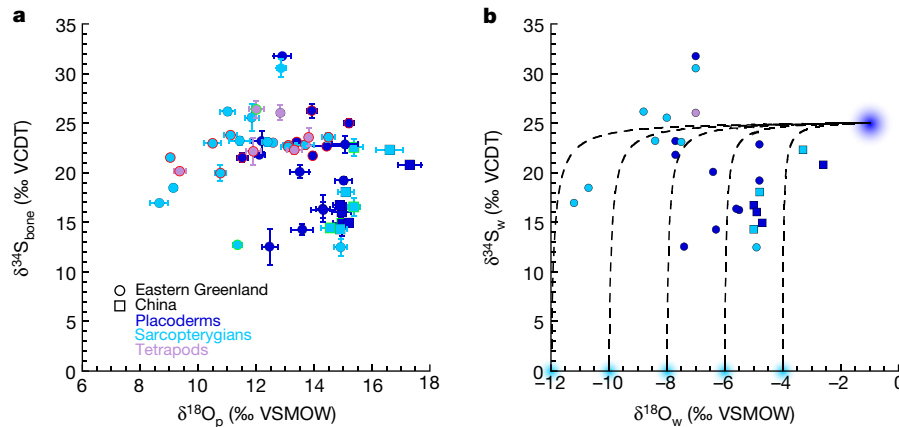
### $\delta^{18}\text{O}_\text{p}$ shows a freshwater influence

Oxygen present in vertebrate bone phosphates derives principally from ingested environmental water<sup>42</sup>. Marine environments have  $\delta^{18}\text{O}$  values that range from  $-2\text{‰}$  to  $+2\text{‰}$ , and in most cases they are  $^{18}\text{O}$ -enriched relative to freshwater, except for freshwater from dry tropical environments<sup>43</sup>. Consequently, marine vertebrates tend to have higher  $\delta^{18}\text{O}_\text{p}$  values than those of freshwater vertebrates. The  $\delta^{18}\text{O}_\text{p}$  values of Devonian vertebrates from East Greenland and China range from  $+8.7\text{‰}$  to  $+15.1\text{‰}$  (mean  $\delta^{18}\text{O}_\text{p} = +12.7\text{‰}$ ;  $1\sigma = 1.8\text{‰}$ ) and  $+14.9\text{‰}$  to  $+17.3\text{‰}$  (mean  $\delta^{18}\text{O}_\text{p} = +15.6\text{‰}$ ;  $1\sigma = 1.0\text{‰}$ ), respectively (Fig. 3a and Supplementary Table 2). These  $\delta^{18}\text{O}_\text{p}$  values are several per mille lower than those previously reported for coeval Devonian conodonts<sup>44</sup>, which are considered to be restricted to marine environments<sup>45</sup> (Fig. 4). The low  $\delta^{18}\text{O}_\text{p}$  values of Devonian vertebrates are therefore not compatible with a pure seawater source, and we therefore interpret them as reflecting a freshwater source of ingested oxygen.

Assuming a mean temperature of  $28^\circ\text{C}$ <sup>44</sup> for Devonian surface waters in East Greenland and China—both of which are located at near-equatorial or tropical palaeolatitudes—we estimated<sup>46</sup> the oxygen isotope composition of environmental waters ( $\delta^{18}\text{O}_\text{w}$ ) in which Devonian vertebrates lived. The  $\delta^{18}\text{O}_\text{w}$  values range from  $-11.2\text{‰}$  to  $-4.8\text{‰}$  (mean  $\delta^{18}\text{O}_\text{w} = -7.2\text{‰}$ ;  $1\sigma = 1.8\text{‰}$ ) and  $-5.0\text{‰}$  to  $-2.6\text{‰}$  (mean  $\delta^{18}\text{O}_\text{w} = -4.3\text{‰}$ ;  $1\sigma = 1.0\text{‰}$ ) for East Greenland and China, respectively (Fig. 3b). Whereas the reconstructed  $\delta^{18}\text{O}_\text{w}$  values for China can be expected for low-latitude freshwater environments, the  $\delta^{18}\text{O}_\text{w}$  values of East Greenland may be considered as low with respect to typical modern tropical or equatorial  $\delta^{18}\text{O}_\text{w}$  values. However, such low values can be observed for modern-day tropical rivers, and have been reconstructed for the drainage system of the Himalayan foreland basin: indeed, the  $\delta^{18}\text{O}_\text{w}$  values of a whole river system can be strongly lowered owing to the catchment of waters having low  $\delta^{18}\text{O}_\text{w}$  values formed at high altitudes and during monsoon rainfalls<sup>47</sup>. East Greenland ORS sediments were deposited in basins related to the Caledonides orogens that probably drained strongly  $^{18}\text{O}$ -depleted meteoric waters formed at high altitudes, hence accounting for the observed  $\delta^{18}\text{O}_\text{p}$  values of Greenland Devonian vertebrates.

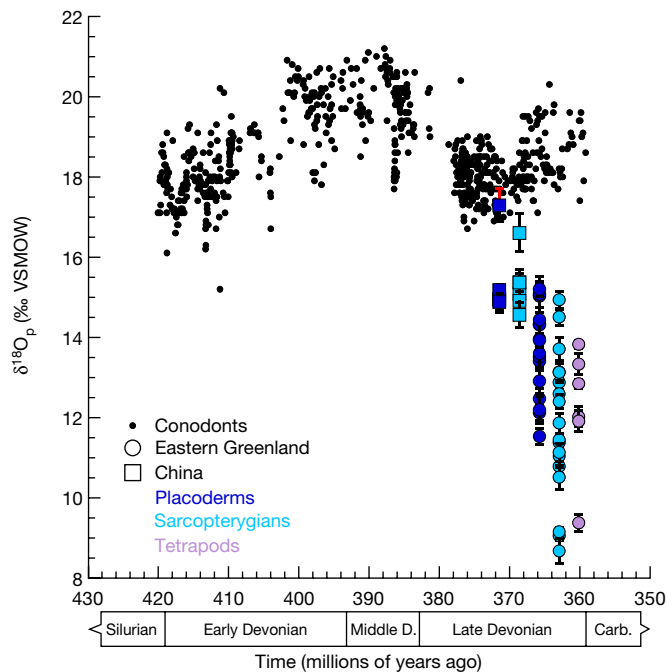
### Palaeontological implications

Overall, the most robust isotopic data demonstrate that Devonian vertebrates from East Greenland and China had isotopic compositions that are indicative of living in a mixture of freshwater and seawater (Fig. 3b). Such an isotopic record is compatible with transitional environments characterized by high-magnitude, rapid changes in salinity



**Fig. 3 | Environmental interpretation.** **a**, Sulfur and oxygen isotope compositions recorded in the mineralized tissues of Devonian vertebrates. Data outlined in red and green correspond to samples that were potentially altered during diagenesis (see Extended Data Fig. 2 and ‘State of preservation of Devonian apatites’ above). **b**, Modelled sulfur and oxygen isotope compositions of the environmental waters of the Devonian vertebrates (See ‘Isotopes mixing in transitional environments’ in Methods). Dashed lines represent the evolution of oxygen and sulfur

(see Methods), such as estuaries or deltas. These results indicate that early tetrapods and their associated vertebrate fauna were most probably euryhaline. The capacity to cope with large changes in salinity is consistent with numerous findings of early tetrapods preserved in Devonian sedimentary deposits that are interpreted as transitional environments<sup>26</sup>.



**Fig. 4 |  $\delta^{18}\text{O}_p$  of Devonian vertebrate bone apatite.**  $\delta^{18}\text{O}_p$  values of Devonian vertebrates are compared to published values of conodonts<sup>44</sup>, which are considered to be restricted to marine environments<sup>45</sup>. The comparison indicates that the  $\delta^{18}\text{O}_p$  values of Devonian vertebrates are not compatible with a strict marine environment, and reveals the presence of freshwater in their living environment. Each dot represents an independent Devonian apatite sample ( $n = 51$ ) and corresponds to the average value of five repeated measurements. Each error bar corresponds to 1 s.d. (Supplementary Table 2). Middle D., Middle Devonian; Carb., Carboniferous.

isotope compositions resulting from a mixing between a marine reservoir (dark blue) and several freshwater reservoirs (pale blue). In **a**, **b**, each dot represents an independent Devonian apatite sample ( $n = 51$ ) and corresponds to the average value of three repeated measurements. Each error bar corresponds to 1 s.d. (Supplementary Table 2).  $\delta^{18}\text{O}_p$  and  $\delta^{18}\text{O}_w$  results are given as variations in parts per mille from the ratio of  $^{18}\text{O}/^{16}\text{O}$  in Vienna Standard Mean Ocean Water (VSMOW).

Consequently, on a broader scale, these results allow for the possibility that many Late Devonian vertebrate species may have been euryhaline and that ORS sediments may have been subjected to greater marine influences than previously thought. This hypothesis may help resolve apparent inconsistencies between the fossil and sedimentary records. For instance, it may explain why *Groenlandaspis*—a placoderm often considered to be a freshwater species—had a global distribution during the Late Devonian<sup>14</sup>. It may also explain the presence of an *Ichthyostega*-like tetrapod in Europe<sup>22</sup>, which was recovered from a sedimentary deposit that was separated from East Greenland by seawater. Finally, it may also explain how early tetrapods had already achieved a worldwide distribution by the Late Devonian<sup>26</sup>, and why several species of tetrapods co-occurred within the same fossil assemblage<sup>25</sup>. Indeed, the mixing of seawater with freshwater in estuaries results in complex dynamic ecosystems, comprising many distinct ecological niches<sup>48</sup> in which species can specialize. For instance, the higher  $\delta^{18}\text{O}_p$  values recorded in the mineralized tissues of placoderms ( $\delta^{18}\text{O}_p$  average in placoderms =  $+13.6 \pm 1.1\text{‰}$  versus  $\delta^{18}\text{O}_p$  average in sarcopterygians and tetrapods =  $+11.8 \pm 1.6\text{‰}$ ;  $P = 0.0007$  (paired  $t$ -test)) may indicate that they were more confined to the seawater end-member than were coeval sarcopterygians and tetrapods, which may have been able to sustain and exploit a greater salinity range (especially towards lower salinities).

Euryhalinity may have predisposed the early tetrapod clade to be able to survive the numerous biotic crises that occurred during the later part of the Devonian<sup>49</sup>, and subsequently to colonize terrestrial habitats from this wide range of aquatic environments. This may explain why their radiation was so successful and how they achieved considerable diversity as early as the Tournaisian age<sup>50</sup>.

## Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0159-2>.

Received: 4 July 2017; Accepted: 15 March 2018;  
Published online: 30 May 2018

- Ahlberg, P. E. *Elginerpeton pancheni* and the earliest tetrapod clade. *Nature* **373**, 420–425 (1995).
- Niedzwiedzki, G., Szrek, P., Narkiewicz, K., Narkiewicz, M. & Ahlberg, P. E. Tetrapod trackways from the early Middle Devonian period of Poland. *Nature* **463**, 43–48 (2010).
- Coates, M. I. & Clack, J. A. Polydactyly in the earliest known tetrapod limbs. *Nature* **347**, 66–69 (1990).

4. Clack, J. A. et al. A uniquely specialized ear in a very early tetrapod. *Nature* **425**, 65–69 (2003).
5. Pierce, S. E., Clack, J. A. & Hutchinson, J. R. Three-dimensional limb joint mobility in the early tetrapod *Ichthyostega*. *Nature* **486**, 523–526 (2012).
6. Agassiz, L. *Monographie des Poissons Fossiles du Vieux Grès Rouge: Ou Système Dévonien (Old Red Sandstone) des Iles Britanniques et de Russie* (Soleure, Neuchatel, 1845).
7. Woodward, A. S. *Catalogue of the Fossil Fishes in the British Museum (Natural History), Part 2* (British Museum (Natural History), London, 1891).
8. Sæve-Söderbergh, G. *Preliminary Note on Devonian Stegocephalians from East Greenland (Meddelelser om Grønland 94)* (1932).
9. McClay, K. R., Norton, M. G., Coney, P. & Davis, G. H. Collapse of the Caledonian orogen and the Old Red Sandstone. *Nature* **323**, 147–149 (1986).
10. Godwin-Austen, R. On the possible extension of the coal-measures beneath the south-eastern part of England. *Q. J. Geol. Soc.* **11**, 533–536 (1855).
11. Barrell, J. The influence of Silurian-Devonian climates on the rise of air-breathing vertebrates. *Proc. Natl Acad. Sci. USA* **2**, 499–504 (1916).
12. Bray, A. A., Potts, W. T. W. & Milner, A. R. The evolution of the terrestrial vertebrates: environmental and physiological considerations. *Phil. Trans. R. Soc. Lond. B* **309**, 289–322 (1985).
13. Schmitz, B., Åberg, G., Werdelin, L., Forey, P. & Bendix-Almgreen, S. E.  $^{87}\text{Sr}/^{86}\text{Sr}$ , Na, F, Sr, and La in skeletal fish debris as a measure of the paleosalinity of fossil-fish habitats. *Geol. Soc. Am. Bull.* **103**, 786–794 (1991).
14. Schultze, H.-P. & Cloutier, R. in *Devonian Fishes and Plants of Miguasha, Québec, Canada* (eds Schultze, H.-P. & Cloutier, R.) 348–368 (Dr Friedrich Pfeil, Munich, 1996).
15. Carr, R. K. & Jackson, G. L. in *Guide to the Geology and Paleontology of the Cleveland Member of the Ohio Shale* (ed. Hannibal, J. T.) Ch. 5 (Ohio Geological Survey Guidebook, Cleveland, 2010).
16. Laurin, M. & Soler-Gijón, R. in *The Terrestrialization Process: Modelling Complex Interactions at the Biosphere–Geosphere Interface* (Geological Society of London Special Publications 339) 151–179 (Geological Society, London, 2010).
17. Campbell, K. S. W. & Bell, M. W. A primitive amphibian from the Late Devonian of New South Wales. *Alcheringa* **1**, 369–381 (1977).
18. Ahlberg, P. E., Luksevics, E. & Lebedev, O. The first tetrapod finds from the Devonian (Upper Famennian) of Latvia. *Phil. Trans. R. Soc. Lond. B* **343**, 303–328 (1994).
19. Lebedev, O. A. & Clack, J. A. Upper Devonian tetrapods from Andreyevka, Tula region, Russia. *Palaeontology* **36**, 721–734 (1993).
20. Daeschler, E. B., Shubin, N. H., Thomson, K. S. & Amaral, W. W. A Devonian tetrapod from North America. *Science* **265**, 639–642 (1994).
21. Zhu, M., Ahlberg, P. E., Zhao, W. & Jia, L. Palaeontology: first Devonian tetrapod from Asia. *Nature* **420**, 760–761 (2002).
22. Clément, G. et al. Palaeogeography: Devonian tetrapod from western Europe. *Nature* **427**, 412–413 (2004).
23. Lebedev, O. A. A new tetrapod *Jakubsonia livnensis* from the Early Famennian (Devonian) of Russia and palaeoecological remarks on the Late Devonian tetrapod habitats. *Acta Univ. Latv.* **679**, 79–98 (2004).
24. Shubin, N. H., Daeschler, E. B. & Coates, M. I. The early evolution of the tetrapod humerus. *Science* **304**, 90–93 (2004).
25. Clack, J. A., Ahlberg, P. E., Blom, H. & Finney, S. M. A new genus of Devonian tetrapod from North-East Greenland, with new information on the lower jaw of *Ichthyostega*. *Palaeontology* **55**, 73–86 (2012).
26. Blicek, A. et al. The biostratigraphical and palaeogeographical framework of the earliest diversification of tetrapods (Late Devonian). *Geol. Soc. Lond. Spec. Publ.* **278**, 219–235 (2007).
27. Nriagu, J. O. et al. in *Stable Isotopes: Natural and Anthropogenic Sulphur in the Environment* (eds Krouse, H. R. & Grinenko, V. A.) 177–265 (John Wiley & Sons, Chichester, 1991).
28. Nehlich, O. The application of sulphur isotope analyses in archaeological research: a review. *Earth-Sci. Rev.* **142**, 1–17 (2015).
29. Nehlich, O., Barrett, J. H. & Richards, M. P. Spatial variability in sulphur isotope values of archaeological and modern cod (*Gadus morhua*). *Rapid Commun. Mass Spectrom.* **27**, 2255–2262 (2013).
30. Goedert, J., Fourel, F., Amiot, R., Simon, L. & Lécuyer, C. High-precision  $^{34}\text{S}/^{32}\text{S}$  measurements in vertebrate biapatites using purge-and-trap elemental analyser/isotope ratio mass spectrometry technology. *Rapid Commun. Mass Spectrom.* **30**, 2002–2008 (2016).
31. Wagman, D. D., Evans, W. H., Parker, V. B., Schumm, R. H. & Halow, I. *The NBS Tables of Chemical Thermodynamic Properties. Selected Values for Inorganic and C<sub>1</sub> and C<sub>2</sub> Organic Substances in SI Units* (National Standard Reference Data System, 1982).
32. Stepańczyk, B., Szostek, K. & Pawlyta, J. The human bone oxygen isotope ratio changes with aging. *Geochronometria* **41**, 147–159 (2014).
33. Szostek, K. Chemical signals and reconstruction of life strategies from ancient human bones and teeth - problems and perspectives. *Anthropol. Rev.* **72**, 3–30 (2009).
34. Iacumin, P., Bocherens, H., Mariotti, A. & Longinelli, A. Oxygen isotope analyses of co-existing carbonate and phosphate in biogenic apatite: a way to monitor diagenetic alteration of bone phosphate? *Earth Planet. Sci. Lett.* **142**, 1–6 (1996).
35. Vennemann, T. W., Hegner, E., Cliff, G. & Benz, G. W. Isotopic composition of recent shark teeth as a proxy for environmental conditions. *Geochim. Cosmochim. Acta* **65**, 1583–1599 (2001).
36. Zazzo, A., Lécuyer, C. & Mariotti, A. Experimentally-controlled carbon and oxygen isotope exchange between biapatites and water under inorganic and microbially-mediated conditions. *Geochim. Cosmochim. Acta* **68**, 1–12 (2004).
37. Brudevold, F. & Soremark, R. in *Structural and Chemical Organization of Teeth Vol. 2* (ed. Miles, A. E. W.) 247–277 (Academic, New York, 1967).
38. Rink, W. J. & Schwarcz, H. P. Tests for diagenesis in tooth enamel: ESR dating signals and carbonate contents. *J. Archaeol. Sci.* **22**, 251–255 (1995).
39. Passey, B. H. et al. Carbon isotope fractionation between diet, breath CO<sub>2</sub>, and biapatite in different mammals. *J. Archaeol. Sci.* **32**, 1459–1470 (2005).
40. Kampschulte, A. & Strauss, H. The sulfur isotopic evolution of Phanerozoic seawater based on the analysis of structurally substituted sulfate in carbonates. *Chem. Geol.* **204**, 255–286 (2004).
41. Fry, B. & Chumchal, M. M. Sulfur stable isotope indicators of residency in estuarine fish. *Limnol. Oceanogr.* **56**, 1563–1576 (2011).
42. Luz, B., Kolodny, Y. & Horowitz, M. Fractionation of oxygen isotopes between mammalian bone-phosphate and environmental drinking water. *Geochim. Cosmochim. Acta* **48**, 1689–1693 (1984).
43. Craig, H. & Gordon, L. I. in *Stable Isotopes in Oceanographic Studies and Paleotemperatures* (ed. Tongiorgio, E.) 9–130 (Consiglio Nazionale delle Ricerche Laboratorio di Geologia Nucleare, Pisa, 1965).
44. Joachimski, M. M. et al. Devonian climate and reef evolution: insights from oxygen isotopes in apatite. *Earth Planet. Sci. Lett.* **284**, 599–609 (2009).
45. Sweet, W. C. & Donoghue, P. C. Conodonts: past, present, future. *J. Paleontol.* **75**, 1174–1184 (2001).
46. Lécuyer, C., Amiot, R., Touzeau, A. & Trotter, J. Calibration of the phosphate  $\delta^{18}\text{O}$  thermometer with carbonate–water oxygen isotope fractionation equations. *Chem. Geol.* **347**, 217–226 (2013).
47. Licht, A. et al. Asian monsoons in a late Eocene greenhouse world. *Nature* **513**, 501–506 (2014).
48. Bouillon, S., Connolly, R. M. & Gillikin, D. P. in *Treatise on Estuarine and Coastal Science* (eds Wolanski, E. et al.) Ch. 7, 143–173 (Elsevier, Amsterdam, 2011).
49. Sallan, L. C. & Coates, M. I. End-Devonian extinction and a bottleneck in the early evolution of modern jawed vertebrates. *Proc. Natl Acad. Sci. USA* **107**, 10131–10135 (2010).
50. Clack, J. A. et al. Phylogenetic and environmental context of a Tournaisian tetrapod fauna. *Nat. Ecol. Evol.* **1**, 002 (2016).

**Acknowledgements** We thank N. Scharff and B. Lindow (Natural History Museum of Denmark) for authorizing the sampling of Devonian material; M. Creuzé des Châtelliers for access to the zoology collections of the Centre de Ressources pour les Sciences de l'Évolution (CERES, FED 4271, Université de Lyon, Université Claude Bernard Lyon 1); O. de Lataillade (Ferme du Ciron), P. François (Pierrelatte), E. Liatout (Maison Liatout) and the fishery (Maison Pupier) for providing present-day material; and L. De Brito. M.Z. was supported by Natural Science Foundation of China (41530102) and Key Research Program of Frontier Sciences of CAS (QYDZJ-SSW-DQC002). This study was supported by the CNRS INSU program Intervie, and the Institut Universitaire de France (C.L.).

**Reviewer information** *Nature* thanks M. Böttcher, O. Nehlich and the other anonymous reviewer(s) for their contribution to the peer review of this work.

**Author contributions** J.G., R.A. and C.L. conceived the project. J.G., F.F., L.S., X.W. and L.C. conducted the isotopic analyses. J.G. and F. A.-G. conducted elemental analyses. J.G. and G.P. conducted the Fourier-transform infrared spectroscopy analyses. G.D. provided access to some present-day material and G.C. and M.Z. provided access to the Devonian material. G.C., J.-S.S. and M.Z. provided palaeontological expertise to shape the project and interpret the data. J.G., R.A. and C.L. wrote the manuscript.

**Competing interests** The authors declare no competing interests.

#### Additional information

**Extended data** is available for this paper at <https://doi.org/10.1038/s41586-018-0159-2>.

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41586-018-0159-2>.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>.

**Correspondence and requests for materials** should be addressed to J.G. or C.L.  
**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## METHODS

No statistical methods were used to predetermine sample size.

**Collection of present-day samples.** All present-day samples analysed in this study are listed and described in Supplementary Table 1. The four red-eared sliders (*Trachemys scripta elegans*; samples Tr-sc-1, Tr-sc-3–5) and the two teeth of the desert crocodile (*Crocodylus suchus*; samples Cr-su-1 and Cr-su-2) were collected at Zoo Lyon, France. Water has been sampled directly from the basins in which these animals are kept. The fire salamander (*Salamandra salamandra*; sample Sal-sal-1) sample comes from the zoology collection of the Centre de Ressources pour les Sciences de l'Évolution (CERES, FED 4271, Villeurbanne, France). Water has been sampled at the location from which the specimen comes. The five marsh frogs (*Pelophylax ridibundus*; samples Pe-ri-1–5) were all collected from a breeding farm in Pierrelatte (Drôme, France), along with the water of their basin and their food. The four trout (*Oncorhynchus mykiss*, *Salvelinus fontinalis* and *Salmo trutta*; samples On-my-2, On-my-3, Sa-f-2 and Sa-tr-2) were all collected from a breeding farm in Allons (Gironde, France), along with the water of their basin and their food. The wels catfish (*Silurus glanis*; sample Si-gl-1) and the common carp (*Cyprinus carpio*; sample Cy-ca-1) were collected in a lake of the Dombes region (Ain, France), exploited by Maison Liatout. Other fish (*Sander lucioperca*, sample Sa-lu-1; *Dicentrarchus labrax*, sample Di-la-1; *Solea solea*, sample So-so-1; *Limanda limanda*, sample Li-li-1; *Gadus morhua*, sample Ga-mo-1; *Oncorhynchus nerka*, sample On-ne-1) were all collected in the fishery Maison Pupier in Lyon.

**Collection of fossil samples.** The 40 Devonian vertebrate samples from East Greenland were selected mainly from unnumbered specimens housed in the palaeontological collection of the Natural History Museum of Denmark, University of Copenhagen. These samples were collected during palaeontological expeditions in the 1930s–1950s, led by the palaeontologists G. Säve-Söderbergh, E. Jarvik and G. Wängsjö. The East Greenland Devonian localities are situated (from north to south) on Gauss Halvø, Ymer Ø, Ella Ø, Traill Ø, Wegener Halvø and Canning Land<sup>51</sup>. Tetrapod fossil remains come from the Upper Devonian Celsius Bjerg group, which consists mainly of siltstones and sandstones. These sediments have previously been interpreted as being primarily deposited under freshwater fluvial conditions<sup>52</sup>. The Celsius Bjerg group is composed of the Agda Dal, Elsa Dal, Aina Dal, Wimans Bjerg, Britta Dal, Stensjö Bjerg and Obrutschew Bjerg formations<sup>53</sup>. Among these formations, tetrapod remains have been reported from only two formations: the Aina Dal and Britta Dal formations, which form—along with the less-fossiliferous Wimans Bjerg formation—the so-called 'Remigolepis series'<sup>53</sup>. We therefore selected mainly samples from the Remigolepis series (34 out of 40, see Supplementary Table 1). Two fossilized biogenic samples come from the 'Phyllolepis series', which groups the Agda and Elsa Dal formations and is, stratigraphically, immediately above the Remigolepis series. Three fossilized biogenic samples come from the 'Asterolepis series', which is stratigraphically lower than the Remigolepis and Phyllolepis series<sup>53</sup>. Based on a microspore analysis, a Famennian age has previously been proposed for the Remigolepis and Phyllolepis series<sup>67</sup>.

The 11 Devonian vertebrate samples from the Chinese Ningxia Hui autonomous region were collected at the Institute of the Vertebrate Paleontology and Paleoanthropology, Beijing, China. These samples come from the Zhongning formation, which has yielded the remains of *Sinostega pani*—which is the first report of a Devonian tetrapod from Asia<sup>21</sup>. The Zhongning formation consists of mainly siltstones, feldspathic quartzose sandstones and arenaceous limestones, all of which have been interpreted as deposited under non-marine conditions<sup>21</sup>. Based on a microspore analysis, a Famennian age has previously been proposed for the Zhongning formation<sup>54</sup>. For each specimen, about 100 mg of bone powder was sampled using a spherical diamond-tipped drill bit.

**Elemental analysis.** Forty-two Devonian apatite samples, and some associated sedimentary matrix samples, were analysed for their elemental contents (Mg, P, S, Ca, Mn, Fe, Cu, Sr and Ba; Supplementary Table 3). Apatite samples, weighing around 50 mg, were dissolved overnight in screw-top Teflon bombs (Saville) using 2 ml of 14 M nitric acid (HNO<sub>3</sub>) at 150 °C. The solutions were rinsed and diluted to 25 ml with distilled water. Each sample solution was then diluted 100 times in 0.5 M HNO<sub>3</sub> containing 2 p.p.b. indium as internal standard. Mg, Mn, Cu, Sr and Ba were analysed using an iCAP Q inductively coupled plasma-mass spectrometer, and P, S, Ca and P were analysed using an iCAP 6000 series inductively coupled plasma-emission spectrometer (UMR 5276, Ens de Lyon).

**Fourier-transform infrared spectroscopy.** Fourier-transform infrared spectroscopy was used to examine the bonding environments of sulfur present in apatite, and the degree of preservation of the mineral structure of samples prepared for the stable isotope analysis. Of the 51 samples, 47 were analysed by Fourier-transform infrared spectroscopy (Supplementary Table 4). Approximately 1 mg of bone powder, added to 40–60 mg of KBr, was ground in an agate mortar, and then compressed to make disks. Absorption infrared spectra were recorded using a Perkin–Elmer GX II FTIR spectrometer (UMR 5306, ILM). The spectra were collected with a spectral resolution of 1 cm<sup>−1</sup> in the 400–4,000 cm<sup>−1</sup> range. Each spectrum was baseline corrected and the absorbance normalized to 1.5.

Following a previously described method<sup>55</sup>, two indicators were calculated to assess the diagenetic alteration of the apatite samples we studied (Supplementary Table 5): the crystallinity index (CI)—given by  $CI = (A_{605} + A_{565})/A_{590}$ —and the carbonate/phosphate indicator ( $CO_3^{2-}/PO_4^{3-}$ ), given by  $CO_3^{2-}/PO_4^{3-} = A_{1,415}/A_{1,035}$ . The crystallinity index represents a combination of the relative sizes of the crystals, as well as the extent to which the atoms in the lattice are ordered. In general, increasing crystallinity index values are related to an increase of crystallinity, as well as an increase in the Ca/P ratio<sup>56</sup>. The  $CO_3^{2-}/PO_4^{3-}$  ratio indicates potential changes in the mineral fraction.  $CO_3^{2-}/PO_4^{3-}$  values below 0.15 indicate the incorporation of secondary apatite with low carbonate content, such as francolite, whereas values above 0.7 indicate the incorporation of secondary calcite<sup>52</sup>.

In a similar way, we also calculated the sulfate/phosphate ratio ( $SO_4^{2-}/PO_4^{3-}$ ; Extended Data Fig. 3) by averaging the following two ratios (Supplementary Table 5):  $SO_4^{2-}/PO_4^{3-} = A_{1,180}/A_{1,035}$  and  $SO_4^{2-}/PO_4^{3-} = A_{635}/A_{605}$ . In these ratios,  $A_x$  is the absorbance (peak intensity) at a given wavenumber ( $x$ ); 565 cm<sup>−1</sup> ( $\delta PO_4^{3-}$ )<sup>55</sup>, 605 cm<sup>−1</sup> ( $\delta PO_4^{3-}$ )<sup>55</sup>, 635 cm<sup>−1</sup> ( $\delta SO_4^{2-}$ )<sup>57,58</sup>, 1,035 cm<sup>−1</sup> ( $\nu PO_4^{3-}$ )<sup>55</sup>, 1,180 cm<sup>−1</sup> ( $\nu SO_4^{2-}$ )<sup>57,58</sup> and 1,415 cm<sup>−1</sup> ( $\nu CO_3^{2-}$ )<sup>55</sup>, as well as the narrowing peak around ~590 cm<sup>−1</sup> that occurs between bands 565 and 605 cm<sup>−1</sup> (ref. <sup>55</sup>).

**$\delta^{18}O_p$  analysis.** All bone apatite samples were treated following a previously described wet chemistry protocol<sup>59</sup>, slightly modified by Lécuyer et al.<sup>60</sup>. This protocol consists of the isolation of phosphate ( $PO_4^{3-}$ ) from apatite as silver phosphate ( $Ag_3PO_4$ ) crystals, using acid dissolution and anion-exchange resin. For each sample, 20–30 mg of enamel powder was dissolved in 2 ml of 2 M HF overnight. The CaF<sub>2</sub> residue was separated by centrifugation and the solution was neutralized by adding 2.2 ml of 2 M KOH. Amberlite anion-exchange resin (2.5 ml) was added to the solution to separate the  $PO_4^{3-}$  ions. After 24 h, the solution was removed and the resin was eluted with 6 ml of 0.5 M NH<sub>4</sub>NO<sub>3</sub>. After 4 h, 0.5 ml of NH<sub>4</sub>OH and 15 ml of an ammoniacal solution of AgNO<sub>3</sub> were added and the samples were placed in a thermostated bath at 70 °C for 7 h, enabling the precipitation of  $Ag_3PO_4$  crystals.

Oxygen isotope compositions were measured using a high-temperature pyrolysis technique involving a VarioPYROcube elemental analyser interfaced in continuous flow mode to an Isoprime isotopic ratio mass spectrometer (the EA-Py-CF-IRMS technique<sup>61,62</sup>, performed at UMR 5276, LGL). For each sample, 5 aliquots of 300 µg of  $Ag_3PO_4$  were mixed with 300 µg of pure graphite powder and loaded in silver foil capsules. Pyrolysis was performed at 1,450 °C. Measurements were calibrated against NBS120c (natural Miocene phosphorite from Florida) and NBS127 (barium sulfate, BaSO<sub>4</sub>;  $\delta^{18}O = 9.3$  ‰). The value of NBS120c was fixed at 21.7 ‰ (VSMOW)<sup>60</sup>. Silver phosphate samples precipitated from standard NBS120c were repeatedly analysed ( $\delta^{18}O_p = 21.8$  ‰;  $1\sigma = 0.3$ ;  $n = 10$ ) along with the silver phosphate samples derived from fossil bioapatites to ensure that no isotopic fractionation occurred during the wet chemistry. The average standard deviation was  $0.25 \pm 0.11$  ‰. s.e.m. Data are reported as  $\delta^{18}O_p$  values versus VSMOW (in ‰).

**$\delta^{13}C_c$  and  $\delta^{18}O_c$  analysis.** Bone apatite samples from East Greenland and China were treated according to a previously published procedure<sup>63</sup>. About 10 mg of apatite powder was washed with a 2% NaOCl solution to remove possible organic matter, followed by a 0.1 M acetic acid solution to remove diagenetic carbonates. The volume of solution/mass of powder ratio was kept constant at 25 µl mg<sup>−1</sup> for both treatments. Each treatment lasted for 24 h and samples were rinsed 5 times with distilled water.

Carbon and oxygen isotope compositions of East Greenland Devonian samples were measured using a Thermo Finnigan Gasbench II coupled in continuous flow to a Finnigan MAT 253 isotope ratio mass spectrometer at the Key Laboratory of Cenozoic Geology and Environment, Chinese Academy of Sciences (Beijing, China). For each pre-treated sample, two aliquots of about 2 mg were reacted with 100% orthophosphoric acid at 72 °C for 1 h under an He atmosphere, according to a previously developed method<sup>64</sup>. Carbon and oxygen isotope compositions of Chinese Devonian samples were measured using a MultiPrep automated preparation system coupled to an isotopic ratio mass spectrometer Isoprime in dual-inlet mode at the Laboratoire de Géologie de Lyon (UMR 5276, Université Claude Bernard Lyon 1). For each pre-treated sample, two aliquots of about 2 mg were automatically reacted with anhydrous oversaturated phosphoric acid at 90 °C for 20 min, according to a previously described method<sup>65</sup>.

The measured carbon and oxygen isotopic compositions were normalized relative to the NBS-19 calcite standard. The normalization incorporates the CO<sub>2</sub>-carbonate acid fractionation factor for calcite. Internal reproducibility for the carbon and oxygen isotopic compositions of apatite carbonate is better than  $\pm 0.1$  ‰ and  $\pm 0.2$  ‰, respectively. Data are reported as  $\delta^{13}C_c$  and  $\delta^{18}O_c$  values versus VPDB and VSMOW respectively.

**$\delta^{34}S$  analysis.** Water samples were filtered on a Millipore system using 0.45-µm acetate cellulose filters. Filtered solutions were then heated to 70 °C and a 5% solution of barium dichlorate was added drop-to-drop to precipitate sulfates as barium

sulfate. Muscles were cleaned with double deionized water and then heated for two days in oven at 70 °C and then crushed into fine powders.

Sulfur isotope compositions were measured using a VarioPYROcube elemental analyser in NCS combustion mode interfaced in continuous-flow mode with an Isoprime 100 isotope ratio mass spectrometer at the platform 'Ecologie Isotopique' of the 'Laboratoire d'Ecologie des Hydrosystèmes Naturels et Anthropisés' (LEHNA, UMR 5023, Villeurbanne, France). Barium sulfates from water samples were analysed by weighing 3 aliquots of 250 µg in tin foil capsules. Measurements were calibrated against the three barium sulfate international standards, NBS127, IAEA-SO-5 and IAEA-SO-6. Muscle samples were analysed by weighing 3 aliquots of 3 mg in tin foil capsules, and for bone samples 3 aliquots of 7 mg of bio-apatite powder were mixed with 20 mg of pure tungsten oxide (WO<sub>3</sub>) powder and loaded in tin foil capsules. Tungsten oxide is a powerful oxidant that ensures the full thermal decomposition of apatite sulfate into sulfur dioxide (SO<sub>2</sub>) gas<sup>30</sup>. Measurements were calibrated against the NBS127 and S1 (silver sulfide, Ag<sub>2</sub>S) international standards. For each analytical run of bone samples, we have also analysed BCR32 samples as a compositional and isotopic standard (S% = 0.72, certified value; δ<sup>34</sup>S = 18.4‰<sup>30,66</sup>) to ensure that analytical conditions were optimal to perform sulfur isotope analyses on samples with low-sulfur content. The standard deviation of δ<sup>34</sup>S measurements was better than 0.3‰. Data are reported as δ<sup>34</sup>S vs. VCDT. The VarioPYROcube elemental analyser was also used to measure the sulfur content of samples (Supplementary Tables 1, 2).

**Isotopes mixing in transitional environments.** In terms of geochemical budget, estuaries can be simply defined by the mixing of two end-member reservoirs of different oxygen and sulfur isotope compositions that are freshwater (δ<sup>18</sup>O<sub>fw</sub>, δ<sup>34</sup>S<sub>fw</sub>) and seawater (δ<sup>18</sup>O<sub>sw</sub>, δ<sup>34</sup>S<sub>sw</sub>). The oxygen isotope composition of the mixing waters (δ<sup>18</sup>O<sub>MIX</sub>) behaves conservatively and is defined by the following equation:

$$\delta^{18}\text{O}_{\text{MIX}} = \frac{(f \times \delta^{18}\text{O}_{\text{fw}} + s \times \delta^{18}\text{O}_{\text{sw}})}{(f + s)} \quad (1)$$

Owing to the higher salinity of seawater, the sulfate concentration of seawater (S<sub>sw</sub>) is 100–1,000<sup>41</sup> times higher than the sulfate concentration of freshwater (S<sub>fw</sub>). *f* and *s* denote the relative fraction of freshwater and seawater, respectively, of the mixing waters. The sulfur isotope composition of the mixing waters (δ<sup>34</sup>S<sub>MIX</sub>) is therefore defined by the following equation:

$$\delta^{34}\text{S}_{\text{MIX}} = \frac{(f \times S_{\text{fw}} \times \delta^{34}\text{S}_{\text{fw}} + s \times S_{\text{sw}} \times \delta^{34}\text{S}_{\text{sw}})}{(f \times S_{\text{fw}} + s \times S_{\text{sw}})} \quad (2)$$

with *f* + *s* = 1. From equations (1) and (2), we can derive the relation between the δ<sup>34</sup>S<sub>MIX</sub> and the δ<sup>18</sup>O<sub>MIX</sub> of mixing waters along an estuarine profile:

$$\delta^{34}\text{S}_{\text{MIX}} = \frac{(\delta^{18}\text{O}_{\text{MIX}} \times (\rho \times \delta^{34}\text{S} - \delta^{34}\text{S}_{\text{sw}}) - \rho \times \delta^{18}\text{O}_{\text{sw}} \times \delta^{34}\text{S}_{\text{fw}} + \delta^{18}\text{O}_{\text{fw}} \times \delta^{34}\text{S}_{\text{sw}})}{(\delta^{18}\text{O}_{\text{MIX}} \times (\rho - 1) - \rho \times \delta^{18}\text{O}_{\text{sw}} + \delta^{18}\text{O}_{\text{fw}})}$$

with

$$\rho = \frac{S_{\text{fw}}}{S_{\text{sw}}}$$

If we consider that S<sub>sw</sub> is 100 times higher than S<sub>fw</sub>, then the mixing of freshwater and marine water results in a typical profile—observed in the modern estuarine environment<sup>41</sup>—in which the δ<sup>34</sup>S<sub>MIX</sub> values rapidly increase at lower salinities to

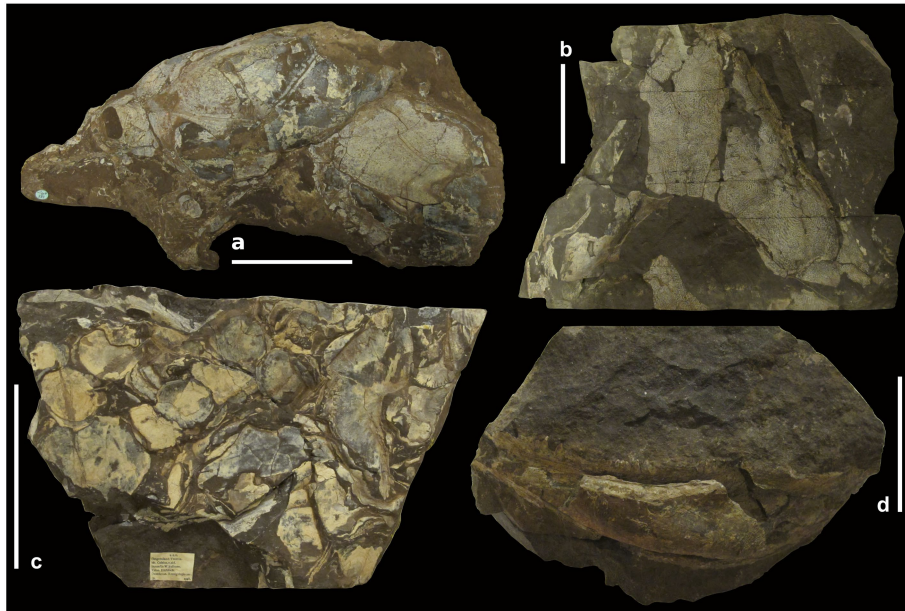
reach the near-constant seawater value (Fig. 3b). Consequently, an animal living in an estuary records in its tissues δ<sup>18</sup>O values that are the result of the linear contribution of the δ<sup>18</sup>O value of both freshwater and seawater end-members, whereas the δ<sup>34</sup>S values are strongly buffered by the δ<sup>34</sup>S value of the seawater end-member.

The δ<sup>18</sup>O<sub>p</sub> and δ<sup>34</sup>S<sub>bone</sub> values we recorded from the Devonian vertebrate bones are consistent with a model of mixing between a δ<sup>18</sup>O-variable freshwater reservoir (δ<sup>18</sup>O<sub>fw</sub> = −4‰ to −12‰; δ<sup>34</sup>S<sub>fw</sub> = 0‰) and a near steady-state seawater reservoir (δ<sup>18</sup>O<sub>sw</sub> = −1‰<sup>44</sup>; δ<sup>34</sup>S<sub>sw</sub> = 25‰<sup>40</sup>) (Fig. 3b).

**Reporting summary.** Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

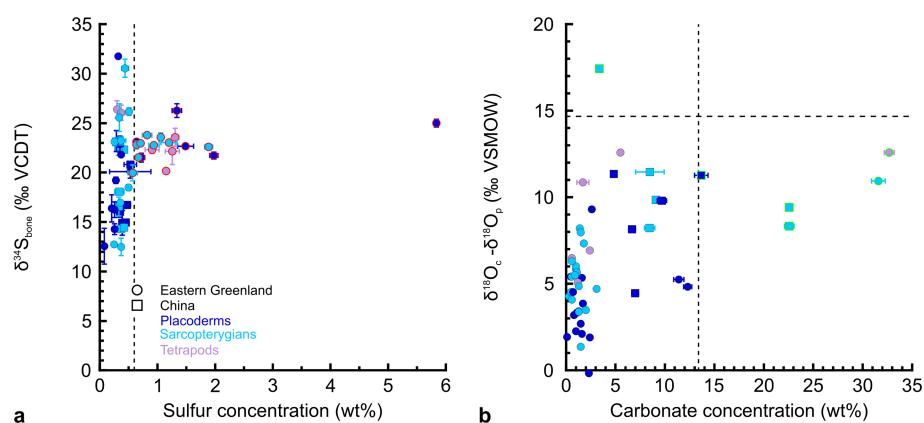
**Data availability statement.** The authors declare that the data supporting the findings of this study are available within the article and its Supplementary Data files (Supplementary Tables 1–5).

- Blom, H., Clack, J. A. & Ahlberg, P. E. Localities, Distribution and Stratigraphical Context of the Late Devonian Tetrapods of East Greenland (Meddelelser om Grønland Geoscience 43) (Danish Polar Center, Copenhagen 2005).
- Olsen, H. & Larsen, R. H. Lithostratigraphy of the continental Devonian sediments in North-East Greenland. *Bull. Grønsl. Geol. Undersøg.* **165**, 1–108 (1993).
- Blom, H., Clack, J. A., Ahlberg, P. E. & Friedman, M. Devonian vertebrates from East Greenland: a review of faunal composition and distribution. *Geodiversitas* **29**, 119–141 (2007).
- Pan, J. et al. *Continental Devonian System of Ningxia and its Biotas* (Geological Publishing House, Beijing, 1987).
- Shemesh, A. Crystallinity and diagenesis of sedimentary apatites. *Geochim. Cosmochim. Acta* **54**, 2433–2438 (1990).
- Termine, J. D. & Posner, A. S. Infrared analysis of rat bone: age dependency of amorphous and crystalline mineral fractions. *Science* **153**, 1523–1525 (1966).
- Nouri, F. Elaboration, caractérisation physico-chimique et étude des propriétés de conduction ionique et de luminescence des fluoroapatites phosphosulfatées dopées à l'euporium. PhD thesis, Univ. of Carthage (2017).
- Nouri, F., Panczer, G., Guyot, Y., Trabelsi-Ayadi, M. & Ternane, R. Synthesis and luminescent properties of Eu<sup>3+</sup>-doped phosphate-sulfate fluorapatites Ca<sub>10-x</sub>Na<sub>x</sub>(PO<sub>4</sub>)<sub>6-x</sub>(SO<sub>4</sub>)<sub>x</sub>F<sub>2</sub>. *J. Lumin.* **192**, 590–594 (2017).
- Crowson, R. A., Showers, W. J., Wright, E. K. & Hoering, T. C. Preparation of phosphate samples for oxygen isotope analysis. *Anal. Chem.* **63**, 2397–2400 (1991).
- Lécuyer, C., Grandjean, P., O'Neil, J. R., Cappetta, H. & Martineau, F. Thermal excursions in the ocean at the Cretaceous–Tertiary boundary (northern Morocco): δ<sup>18</sup>O record of phosphatic fish debris. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **105**, 235–243 (1993).
- Lécuyer, C. et al. High-precision determination of <sup>18</sup>O/<sup>16</sup>O ratios of silver phosphate by EA-pyrolysis-IRMS continuous flow technique. *J. Mass Spectrom.* **42**, 36–41 (2007).
- Fourel, F. et al. <sup>18</sup>O/<sup>16</sup>O ratio measurements of inorganic and organic materials by elemental analysis-pyrolysis-isotope ratio mass spectrometry continuous-flow techniques. *Rapid Commun. Mass Spectrom.* **25**, 2691–2696 (2011).
- Koch, P. L., Tuross, N. & Fogel, M. L. The effects of sample treatment and diagenesis on the isotopic integrity of carbonate in biogenic hydroxylapatite. *J. Archaeol. Sci.* **24**, 417–429 (1997).
- Spötl, C. & Vennemann, T. W. Continuous-flow isotope ratio mass spectrometric analysis of carbonate minerals. *Rapid Commun. Mass Spectrom.* **17**, 1004–1006 (2003).
- Swart, P. K., Burns, S. J. & Leder, J. J. Fractionation of the stable isotopes of oxygen and carbon in carbon dioxide during the reaction of calcite with phosphoric acid as a function of temperature and technique. *Chem. Geol. Isot. Geosci. Sect.* **86**, 89–96 (1991).
- Fourel, F., Martineau, F., Seris, M. & Lécuyer, C. Measurement of <sup>34</sup>S/<sup>32</sup>S ratios of NBS 120c and BCR 32 phosphorites using purge and trap EA-IRMS technology. *Geostand. Geoanal. Res.* **39**, 47–53 (2015).
- Marshall, J. E. A., Astin, T. R. & Clack, J. A. East Greenland tetrapods are Devonian in age. *Geology* **27**, 637–640 (1999).



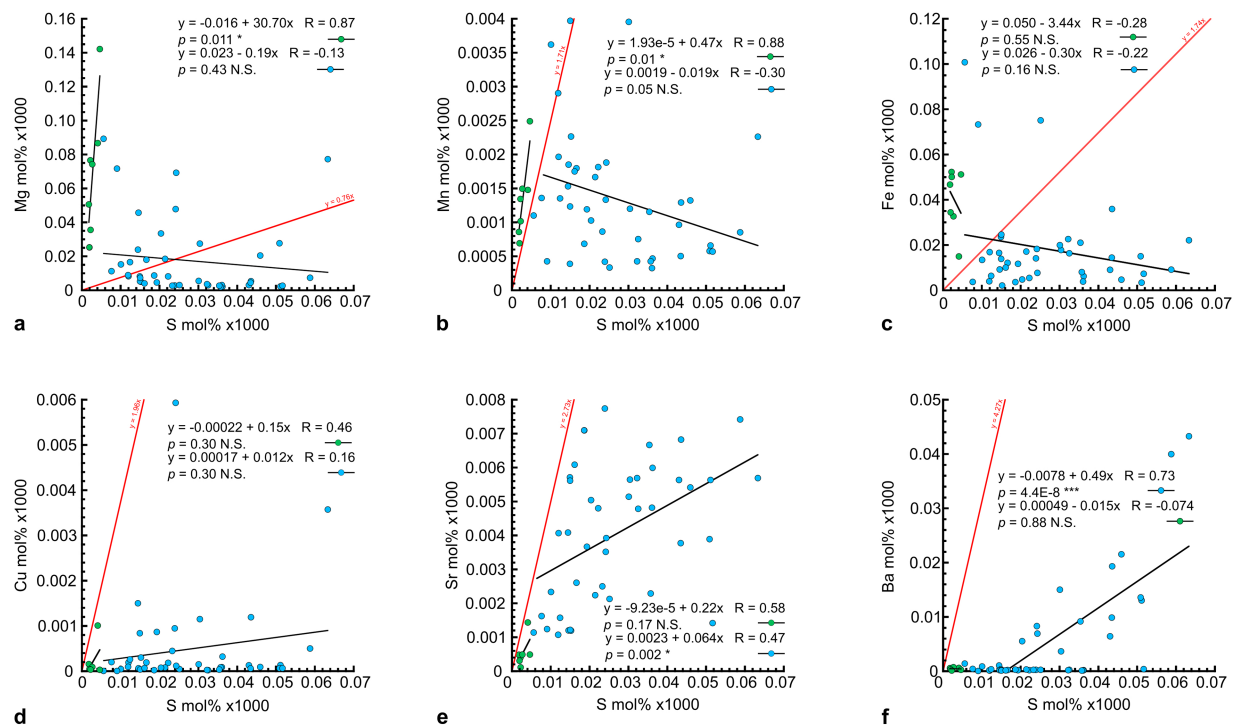
**Extended Data Fig. 1 | Late Devonian vertebrate samples from East Greenland.** **a**, Subcomplete skull of the placoderm *Remigolepis acuta* (4-117; numbers in parentheses refer to sample numbers, reported in Supplementary Table 2). **b**, Subcomplete skull of the sarcopterygian

*Eusthenodon* sp. (27-1151) in dorsal view. **c**, Dermal scales of the sarcopterygians *Holoptychius* sp. (30-1186). **d**, Partial hemimandible of an indeterminate member of Tetrapoda (42-1375). Scale bars, 10 cm.



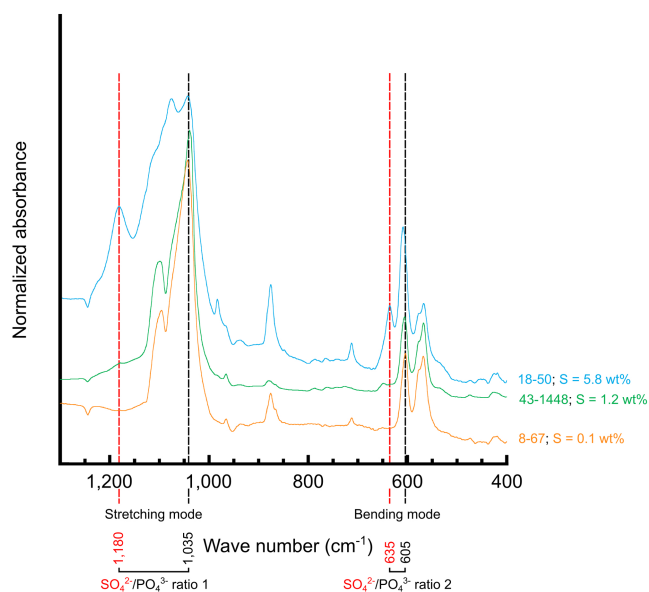
**Extended Data Fig. 2 | Filtering of stable isotope data. a.**  $\delta^{34}\text{S}_{\text{bone}}$  of Devonian tetrapod apatites plotted as a function of sulfur concentration. Samples for which sulfur concentrations are higher than 0.6% (outlined in red) are considered to have been potentially affected by diagenetic alteration. **b.** Covariation of  $\delta^{18}\text{O}_p$  and  $\delta^{18}\text{O}_c$  as a function of carbonate concentration. Samples with both  $\delta^{18}\text{O}_p - \delta^{18}\text{O}_c$  differences higher than

14.7‰ and carbonate content higher than 13.4% (outlined in green) are considered to have been potentially affected by diagenetic alteration. (See also Supplementary Table 2). In **a**, **b**, each dot represents an independent fossil sample ( $n = 51$ ) and corresponds to the average value of three repeated measurements. Each error bar corresponds to 1 s.d. (Supplementary Table 2).

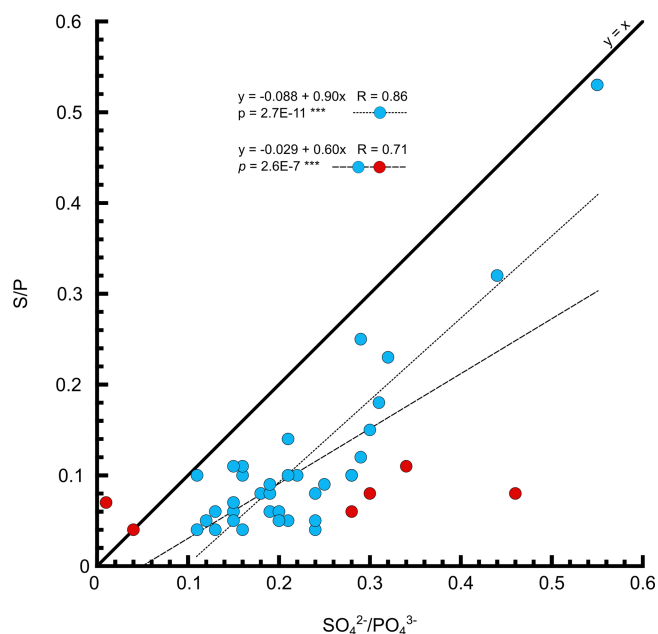


**Extended Data Fig. 3 | Elemental analysis. a–f,** Sulfur content (mol%  $\times$  1,000) of apatites from Devonian tetrapod is graphically represented against various elements to test possible associations as secondary precipitated mineral phases in fossil bioapatites. The elements tested were magnesium (Mg; **a**), manganese (Mn; **b**), iron (Fe; **c**), copper (Cu; **d**), strontium (Sr; **e**) and barium (Ba; **f**). In **a–f**, each dot represents an independent Devonian apatite sample (light blue,  $n = 42$ ; Supplementary

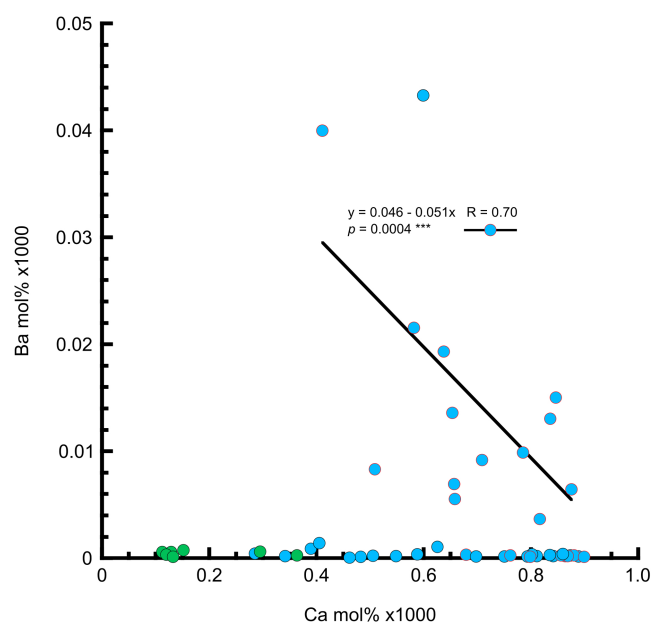
Table 3) or associated matrix sample (light green,  $n = 7$ ; Supplementary Table 3). A 0.6 wt% of sulfur is equal to 0.019 mol%  $\times$  1,000. Red straight lines correspond to calculated stoichiometric lines between the two elements. Black straight lines have been calculated using Pearson's correlation test between the two elements for Devonian apatite samples and associated matrices. N.S., no significant correlation.



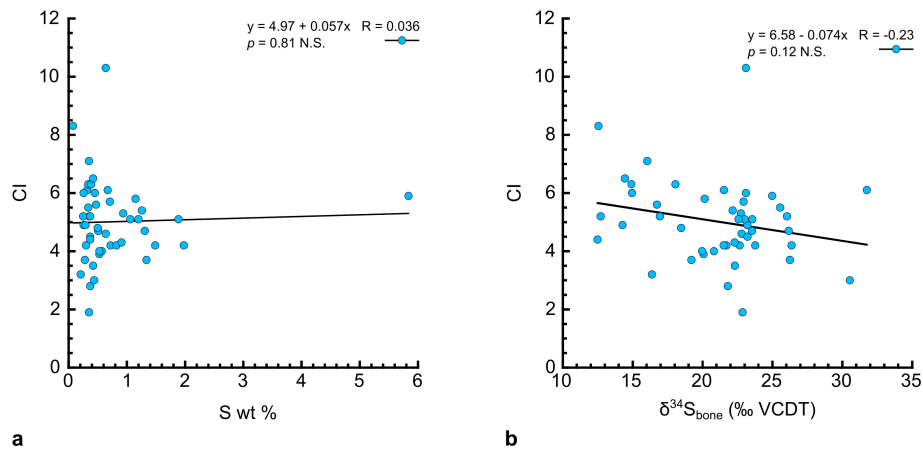
**Extended Data Fig. 4 | Fourier-transform infrared spectroscopy analysis.** Spectra of three representative fossilized biogenic apatites of Devonian vertebrates with various sulfur contents (wt% as measured by the VarioPyrocube elemental analyser). The [PO<sub>4</sub>] and [SO<sub>4</sub>] absorption bands used to calculate the SO<sub>4</sub><sup>2-</sup>/PO<sub>4</sub><sup>3-</sup> ratios are indicated in black and red dashed lines, respectively. Each spectrum corresponds to a single measurement (Supplementary Table 4).



**Extended Data Fig. 5 |  $\text{SO}_4^{2-}/\text{PO}_4^{3-}$  versus S/P ratios.**  $\text{SO}_4^{2-}/\text{PO}_4^{3-}$  ratios of apatites from Devonian tetrapods calculated from the infrared spectra are graphically represented against the S/P ratios calculated from elemental analysis. The two calculated regression lines (dashed) are both highly significant (Pearson's correlation test;  $n = 42$  and  $n = 36$ , respectively) and show a slope close to 1, which indicates that the majority of the sulfur present in apatite is in the form of sulfate that is structurally substituted for the phosphate in the crystal lattice of the apatite. Each dot represents an independent Devonian apatite sample ( $n = 42$ ; Supplementary Tables 3, 5).

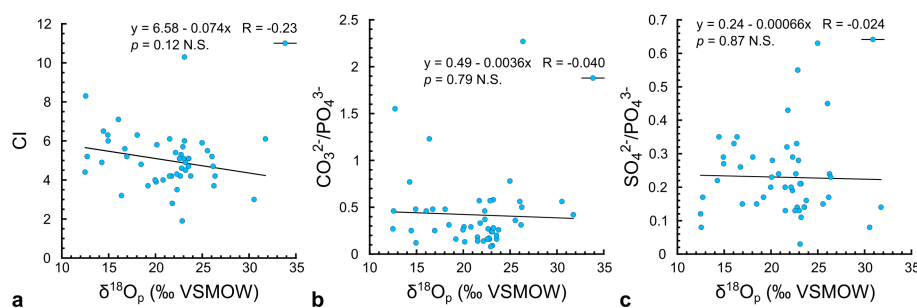


**Extended Data Fig. 6 | Calcium versus barium.** The apatite samples of Devonian tetrapods with a sulfur content over  $0.019 \text{ mol\%} \times 1,000$  (Supplementary Table 3) have been outlined with a red circle. These samples display a significant negative correlation for their calcium and barium elements (Pearson's correlation test;  $n = 21$ ). This observation suggests that calcium is substituted with barium during the recrystallization process. In the same manner, some sulfates may also substitute in for the phosphates within the apatite lattice. Each dot represents an independent Devonian apatite sample (light blue,  $n = 42$ ; Supplementary Table 3) or matrix sample (light green,  $n = 7$ ; Supplementary Table 3).



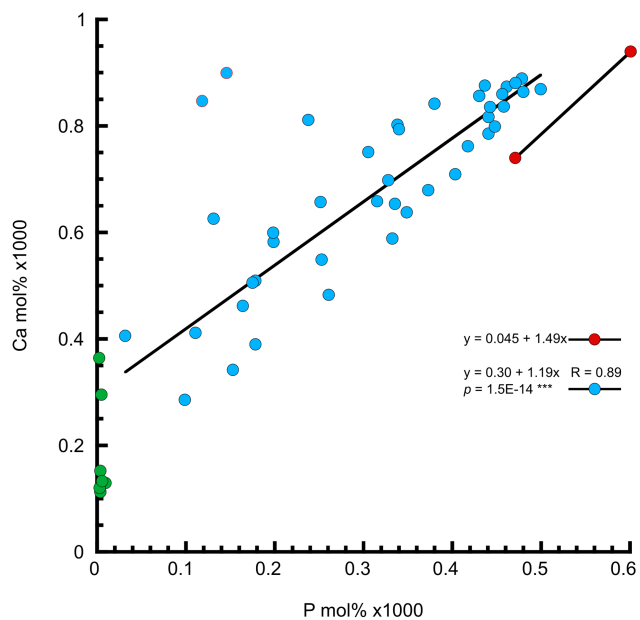
**Extended Data Fig. 7 | Sulfur preservation. a, b,** Sulfur content (a) and isotope composition (b) of apatites from Devonian tetrapods are graphically represented against the crystallinity index (CI). None of them displays a significant correlation (Pearson's correlation test;  $n = 47$ ) with the crystallinity index, which indicates that recrystallization processes

were not systematically associated with sulfate incorporation by elemental substitution and alteration of the sulfur isotope compositions. In a, b, each dot represents an independent Devonian apatite sample ( $n = 42$ ; Supplementary Tables 2, 5).



**Extended Data Fig. 8 | Oxygen preservation.** **a–c**, The oxygen isotope composition of apatites from Devonian tetrapods is graphically represented against the crystallinity index (CI; **a**), the  $\text{CO}_3^{2-}/\text{PO}_4^{3-}$  ratio (**b**) and the  $\text{SO}_4^{2-}/\text{PO}_4^{3-}$  ratio (**c**). None of these indicators displays a significant correlation (Pearson's correlation test;  $n = 47$ ) with the

oxygen isotope composition, thus arguing in favour of at least a partial preservation of the pristine oxygen isotope composition. In **a–c**, each dot represents an independent Devonian apatite sample ( $n = 47$ ; Supplementary Tables 2, 5).



**Extended Data Fig. 9 | Phosphorus versus calcium.** The phosphorus (P) content of apatites from Devonian tetrapods is graphically represented against the calcium (Ca) content. On the whole, the Ca and P contents of Devonian samples are significantly correlated (Pearson's correlation test;  $n = 42$ ) with a slope of 1.19, close to that defined by the NIST 1400 and NIST 1486, both of which are modern bones (red dots;  $\text{Ca/P} = 1.49$ ). These results indicate that the loss of phosphate during recrystallization was almost stoichiometric for all samples. Each dot represents an independent Devonian apatite sample (light blue,  $n = 42$ ; Supplementary Table 3), a matrix sample (light green,  $n = 7$ ; Supplementary Table 3) or an international standard sample (red,  $n = 2$ ; Supplementary Table 3).

# Genomic atlas of the human plasma proteome

Benjamin B. Sun<sup>1,22</sup>, Joseph C. Maranville<sup>2,20,22</sup>, James E. Peters<sup>1,3,22</sup>, David Stacey<sup>1</sup>, James R. Staley<sup>1</sup>, James Blackshaw<sup>1</sup>, Stephen Burgess<sup>1,4</sup>, Tao Jiang<sup>1</sup>, Ellie Paige<sup>1,5</sup>, Praveen Surendran<sup>1</sup>, Clare Oliver-Williams<sup>1,6</sup>, Mihir A. Kamat<sup>1</sup>, Bram P. Prins<sup>1</sup>, Sheri K. Wilcox<sup>7</sup>, Erik S. Zimmerman<sup>7</sup>, An Chi<sup>2</sup>, Narinder Bansal<sup>1,8</sup>, Sarah L. Spain<sup>9</sup>, Angela M. Wood<sup>1</sup>, Nicholas W. Morrell<sup>3,10</sup>, John R. Bradley<sup>11</sup>, Nebojsa Janjic<sup>7</sup>, David J. Roberts<sup>12,13</sup>, Willem H. Ouwehand<sup>3,14,15,16,17</sup>, John A. Todd<sup>18</sup>, Nicole Soranzo<sup>3,14,16,17</sup>, Karsten Suhre<sup>19</sup>, Dirk S. Paul<sup>1</sup>, Caroline S. Fox<sup>2</sup>, Robert M. Plenge<sup>2,20</sup>, John Danesh<sup>1,3,16,17\*</sup>, Heiko Runz<sup>2,21,23</sup> & Adam S. Butterworth<sup>1,17,23\*</sup>

**Although plasma proteins have important roles in biological processes and are the direct targets of many drugs, the genetic factors that control inter-individual variation in plasma protein levels are not well understood. Here we characterize the genetic architecture of the human plasma proteome in healthy blood donors from the INTERVAL study. We identify 1,927 genetic associations with 1,478 proteins, a fourfold increase on existing knowledge, including *trans* associations for 1,104 proteins. To understand the consequences of perturbations in plasma protein levels, we apply an integrated approach that links genetic variation with biological pathway, disease, and drug databases. We show that protein quantitative trait loci overlap with gene expression quantitative trait loci, as well as with disease-associated loci, and find evidence that protein biomarkers have causal roles in disease using Mendelian randomization analysis. By linking genetic factors to diseases via specific proteins, our analyses highlight potential therapeutic targets, opportunities for matching existing drugs with new disease indications, and potential safety concerns for drugs under development.**

Plasma proteins have key roles in various biological processes, including signalling, transport, growth, repair, and defence against infection. These proteins are frequently dysregulated in disease and are important drug targets. Identifying factors that determine inter-individual protein variability should, therefore, furnish biological and medical insights<sup>1</sup>. Despite evidence for the heritability of plasma protein abundance<sup>2</sup>, however, systematic assessment of how genetic variation influences plasma protein levels has been limited<sup>3–5</sup>. Studies have examined intracellular protein quantitative trait loci (pQTLs)<sup>6,7</sup>, but these studies have tended to be small and involved cell lines rather than primary human tissues.

Here we create and interrogate a genetic atlas of the human plasma proteome, using an expanded version of an aptamer-based multiplex protein assay (SOMAscan)<sup>8</sup> to quantify 3,622 plasma proteins in 3,301 healthy participants from the INTERVAL study, a genomic biorepository of 50,000 blood donors from 25 centres across England recruited into a randomized trial of blood donation frequency<sup>9,10</sup>. We identify 1,927 genotype–protein associations (pQTLs), including *trans*-associated loci for 1,104 proteins, providing new understanding of the genetic control of protein regulation. Eighty-eight pQTLs overlap with disease susceptibility loci, suggesting the molecular effects of disease-associated variants. Using the principle of Mendelian randomization<sup>11</sup>, we find

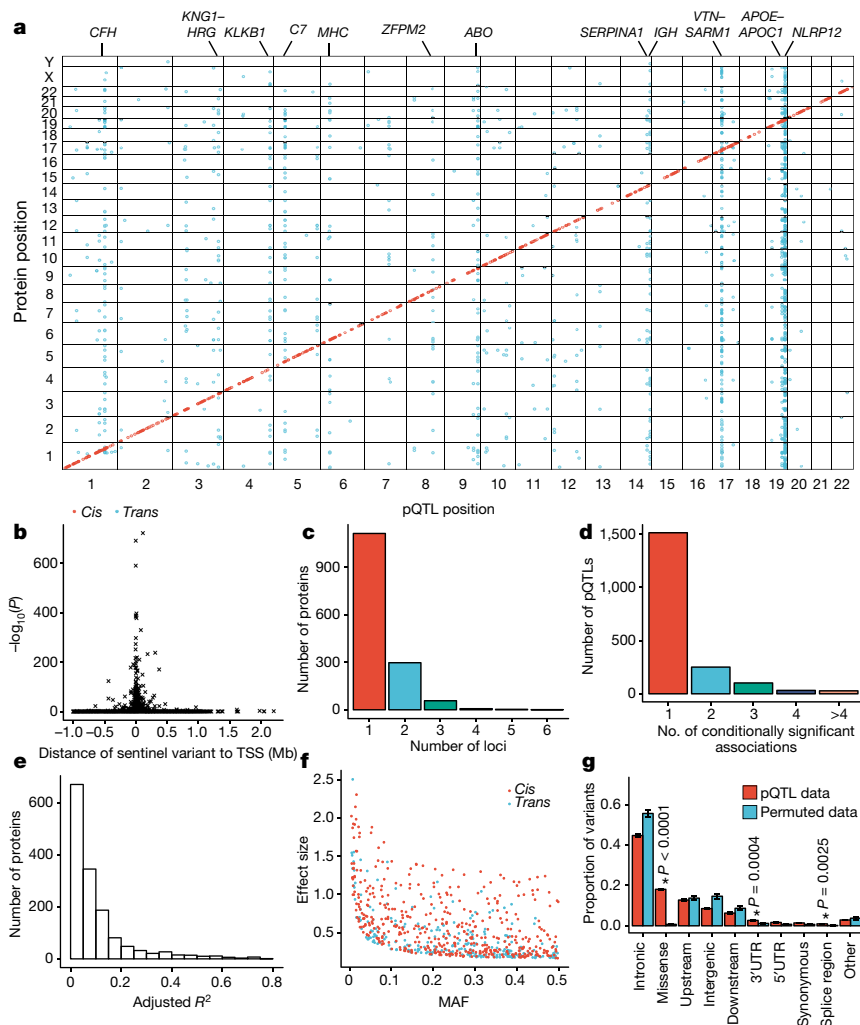
evidence to support causal roles in disease for several protein pathways, and cross-reference our data with disease and drug databases to highlight potential therapeutic targets.

## Genetic architecture of the plasma proteome

We performed genome-wide testing of 10.6 million imputed autosomal variants against levels of 2,994 plasma proteins in 3,301 individuals of European descent (Methods, Extended Data Fig. 1). We demonstrated the robustness of protein measurements in several ways (Supplementary Note, Extended Data Fig. 2), including: highly consistent measurements in replicate samples; temporal consistency of protein levels within individuals over two years (Extended Data Fig. 3b); and replication of known associations with non-genetic factors (Supplementary Tables 1, 2). To assess potential off-target cross-reactivity, we tested 920 aptamers (SOMAmers) for detection of proteins with at least 40% sequence homology to the target protein (Methods). Although 126 (14%) SOMAmers showed comparable binding with a homologous protein (Supplementary Table 3), nearly half of these were binding to alternative forms of the same protein.

We found 1,927 significant ( $P < 1.5 \times 10^{-11}$ ) associations between 1,478 proteins and 764 genomic regions (Fig. 1a, Supplementary Table 4, Supplementary Fig. 1, Supplementary Note Table 1), with

<sup>1</sup>MRC/BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. <sup>2</sup>MRL, Merck & Co., Inc., Kenilworth, NJ, USA. <sup>3</sup>British Heart Foundation Cambridge Centre of Excellence, Division of Cardiovascular Medicine, Addenbrooke's Hospital, Cambridge, UK. <sup>4</sup>MRC Biostatistics Unit, University of Cambridge, Cambridge, UK. <sup>5</sup>National Centre for Epidemiology and Population Health, The Australian National University, Canberra, Australian Capital Territory, Australia. <sup>6</sup>Homerton College, Cambridge, UK. <sup>7</sup>SomaLogic Inc, Boulder, CO, USA. <sup>8</sup>Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK. <sup>9</sup>Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK. <sup>10</sup>Division of Respiratory Medicine, Department of Medicine, University of Cambridge, Cambridge, UK. <sup>11</sup>NIHR Cambridge Biomedical Research Centre/BioResource, Cambridge University Hospitals, Cambridge, UK. <sup>12</sup>National Health Service (NHS) Blood and Transplant and Radcliffe Department of Medicine, NIHR Oxford Biomedical Research Centre, University of Oxford, John Radcliffe Hospital, Oxford, UK. <sup>13</sup>BRC Haematology Theme and Department of Haematology, Churchill Hospital, Oxford, UK. <sup>14</sup>Department of Haematology, University of Cambridge, Cambridge Biomedical Campus, Cambridge, UK. <sup>15</sup>National Health Service (NHS) Blood and Transplant, Cambridge Biomedical Campus, Cambridge, UK. <sup>16</sup>Department of Human Genetics, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK. <sup>17</sup>NIHR Blood and Transplant Research Unit in Donor Health and Genomics, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. <sup>18</sup>JDRF/Wellcome Trust Diabetes and Inflammation Laboratory, Wellcome Trust Centre for Human Genetics, Nuffield Department of Medicine, NIHR Oxford Biomedical Research Centre, University of Oxford, Oxford, UK. <sup>19</sup>Department of Physiology and Biophysics, Weill Cornell Medicine–Qatar, Doha, Qatar. <sup>20</sup>Present address: Celgene Inc., Cambridge, MA, USA. <sup>21</sup>Present address: Biogen Inc., Cambridge, MA, USA. <sup>22</sup>These authors contributed equally: Benjamin B. Sun, Joseph C. Maranville, James E. Peters. <sup>23</sup>These authors jointly supervised this work: Heiko Runz, Adam S. Butterworth. \*e-mail: jd292@medschl.cam.ac.uk; asb38@medschl.cam.ac.uk



**Fig. 1 | The genetic architecture of plasma protein levels.**  $n = 3,301$  participants. **a**, Genomic locations of pQTLs. Red, *cis*; blue, *trans*. The  $x$ - and  $y$ -axes indicate the positions of the sentinel variant and the gene encoding the associated protein, respectively. Highly pleiotropic genomic regions are annotated. **b**, Significance of *cis* associations (linear regression) versus distance of sentinel variant from TSS. **c**, Number of significantly associated loci per protein. **d**, Number of conditionally

significant associations within each pQTL. **e**, Histogram of variance explained by conditionally significant variants. **f**, Effect size versus MAF. **g**, Distributions of the predicted functional annotation classes of sentinel pQTL variants versus null sets of variants from permutation. Bar height represents the mean proportion of variants within each class and error bars reflect one s.d. from the mean. \*Significant enrichment (permutation test, Bonferroni-corrected threshold,  $P < 0.005$ ).

89% of these pQTLs being previously unreported. Of the 764 associated regions, 502 (66%) had local-acting (*cis*) associations only, 228 (30%) *trans* only, and 34 (4%) both *cis* and *trans* (Supplementary Note Table 1). Of the *cis* pQTL sentinel variants, 95% and 87% were located within 200 kb and 100 kb, respectively, of the relevant gene's canonical transcription start site (TSS), and 44% were within the gene itself. The  $P$  values for *cis* associations increased with distance from the TSS (Fig. 1b), mirroring findings for gene expression QTLs (eQTLs)<sup>12</sup>. Of proteins with a significant pQTL, 88% had either *cis* ( $n = 374$ ) or *trans* ( $n = 925$ ) associations only, and 12% ( $n = 179$ ) had both (Supplementary Note Table 1). The majority of significantly associated proteins (75%;  $n = 1,113$ ) had a single pQTL, while 20% had two and 5% had more than two (Fig. 1c). To detect multiple independent associations at the same locus, we used stepwise conditional analysis, identifying 2,658 conditionally significant associations (Supplementary Table 5). Of the 1,927 pQTLs, 414 (21%) had multiple conditionally significant associations (Fig. 1d), of which 255 were *cis*.

We tested replication of 163 pQTLs in 4,998 individuals using an alternative protein assay (Olink, see Methods)<sup>13</sup>. Effect-size estimates were strongly correlated between the SOMAscan and Olink platforms ( $r = 0.83$ ; Extended Data Fig. 3c). One-hundred and six out of one-hundred and sixty-three (65% overall; 81% *cis*,

52% *trans*) pQTLs were replicated after Bonferroni correction (Supplementary Tables 4, 6). The lower replication rate of *trans* associations may reflect various factors, including differences between protein assays (for example, detection of free versus complexed proteins, Extended Data Fig. 4) and the higher 'biological prior' for *cis* associations.

Of 1,927 pQTLs, 549 (28%) were *cis*-acting (Supplementary Table 4). Genetic variants that change protein structure may result in apparent *cis* pQTLs owing to altered aptamer binding rather than true quantitative differences in protein levels. We found evidence against such artefactual associations for 371 (68%) *cis* pQTLs (Methods, Supplementary Tables 4, 7, 8). The results were materially unchanged when we repeated downstream analyses but excluded pQTLs without evidence against binding effects.

The median variation in protein levels explained by pQTLs was 5.8% (interquartile range: 2.6–12.4%, Fig. 1e). For 193 proteins, genetic variants explained more than 20% of the variation. There was a strong inverse relationship between effect size and minor allele frequency (MAF) (Fig. 1f), consistent with previous genome-wide association studies (GWAS) of quantitative traits<sup>7,10,14</sup>. We found 23 and 208 associations with rare (MAF < 1%) and low-frequency (MAF 1–5%) variants, respectively (Supplementary Table 4). Of the 36 strongest associations

(per-allele effect size  $>1.5$  standard deviation (s.d.)), 29 were with rare or low-frequency variants.

Both *cis* and *trans* pQTLs were strongly enriched for missense variants ( $P < 0.0001$ ) and for location in 3' untranslated ( $P = 0.0025$ ) or splice sites ( $P = 0.0004$ ) (Fig. 1g, Extended Data Fig. 5a). We found at least threefold enrichment ( $P < 5 \times 10^{-5}$ ) of pQTLs at features indicative of transcriptional activation in blood cells and at hepatocyte regulatory elements, consistent with the role of the liver in protein synthesis and secretion (Methods, Extended Data Fig. 6, Supplementary Table 9).

### Overlap of eQTLs and pQTLs

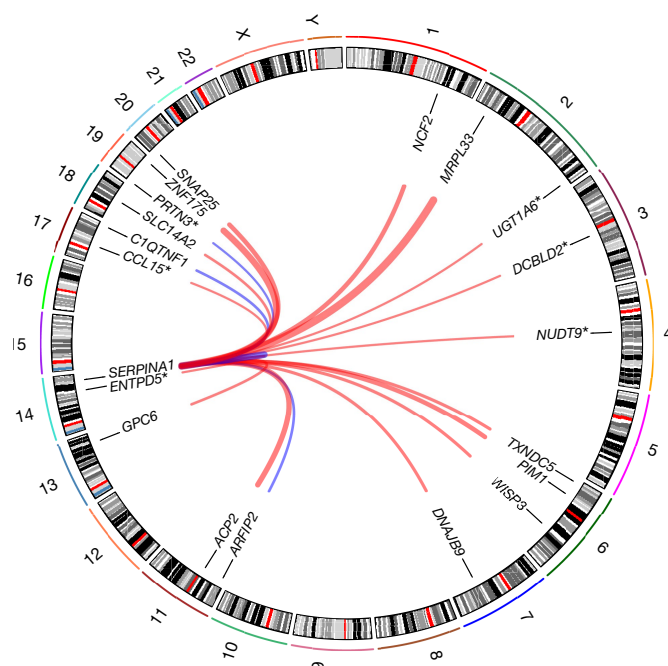
To help evaluate the extent to which genetic associations with plasma protein levels are driven by effects at the transcriptional level rather than other mechanisms (for example, altered protein clearance or secretion), we cross-referenced our *cis* pQTLs with previous eQTL studies (Supplementary Table 10), initially defining overlap between an eQTL and pQTL as high linkage disequilibrium (LD) ( $r^2 \geq 0.8$ ) between the lead pQTL and eQTL variants. Forty per cent ( $n = 224$ ) of *cis* pQTLs were eQTLs for the same gene in one or more tissue or cell type (Supplementary Table 8). The greatest overlaps were in whole blood ( $n = 117$ ), liver ( $n = 70$ ) and lymphoblastoid cell lines (LCLs) ( $n = 52$ ), consistent with biological expectation, but also probably driven by the larger eQTL study sample sizes for these cell types. To investigate whether the same causal variant was likely to underlie overlapping eQTLs and pQTLs, we performed colocalization testing (see Methods). Of 228 pQTLs outside the human leukocyte antigen (HLA) region for which testing was possible, colocalization in one or more tissue or cell type was highly likely (posterior probability (PP)  $> 0.8$ ) for 179 (78.5%) and the most likely explanation (PP  $> 0.5$ ) for 197 (86.4%) (Supplementary Table 8). *cis* pQTLs were significantly enriched for eQTLs for the corresponding gene ( $P < 0.0001$ ) (Methods, Supplementary Table 11). To address the converse (that is, to what extent are eQTLs also pQTLs), we selected well-powered eQTL studies in relevant tissues (whole blood, LCLs, liver and monocytes<sup>15–18</sup>). Of the strongest *cis* eQTLs ( $P < 1.5 \times 10^{-11}$ ) in whole blood, LCLs, liver and monocytes, 12.2%, 21.3%, 14.8% and 14.7%, respectively, were plasma *cis* pQTLs.

Comparisons between eQTL and pQTL studies have inherent limitations, including differences in the tissues, sample sizes and technological platforms used. Moreover, plasma protein levels may not reflect levels within tissues or cells. Nevertheless, our data suggest that genetic effects on plasma protein abundance are often, but not exclusively, driven by regulation of mRNA. *cis* pQTLs without corresponding *cis* eQTLs may reflect genetic effects on processes other than transcription, including protein degradation, binding, secretion, or clearance from circulation.

### *trans* pQTLs identify pathways to disease

Of the 764 protein-associated regions, 262 had *trans* associations with 1,104 proteins (Supplementary Tables 4, 12). There was no enrichment of cross-reactivity in SOMAmers with a *trans* pQTL versus those without (Supplementary Note). We replicated known *trans* associations, including *TMPRSS6* with transferrin receptor protein 1<sup>19</sup> and *SORT1* with granulins<sup>20</sup>, and identified several novel and biologically plausible *trans* associations (Supplementary Table 13), including known or presumed ligand–receptor pairs (for example, the *CD320* locus, encoding the transcobalamin receptor, was associated with transcobalamin-2 levels).

Most *trans* loci (82%) were associated with fewer than four proteins, but twelve ‘hotspot’ regions were associated with more than twenty (Fig. 1a, Extended Data Fig. 5b), including well-known pleiotropic loci (for example, *ABO*, *CFH*, *APOE* and *KLKB1*) and loci associated with many correlated proteins (for example, the *ZFPM2* locus, which encodes the transcription factor FOG2). Similar pleiotropy at these loci has been seen in other plasma pQTL studies<sup>3–5</sup>, albeit with fewer proteins owing to limited assay breadth. A missense variant (rs28929474:T) in *SERPINA1* was associated with 13 proteins at  $P < 1.5 \times 10^{-11}$  and

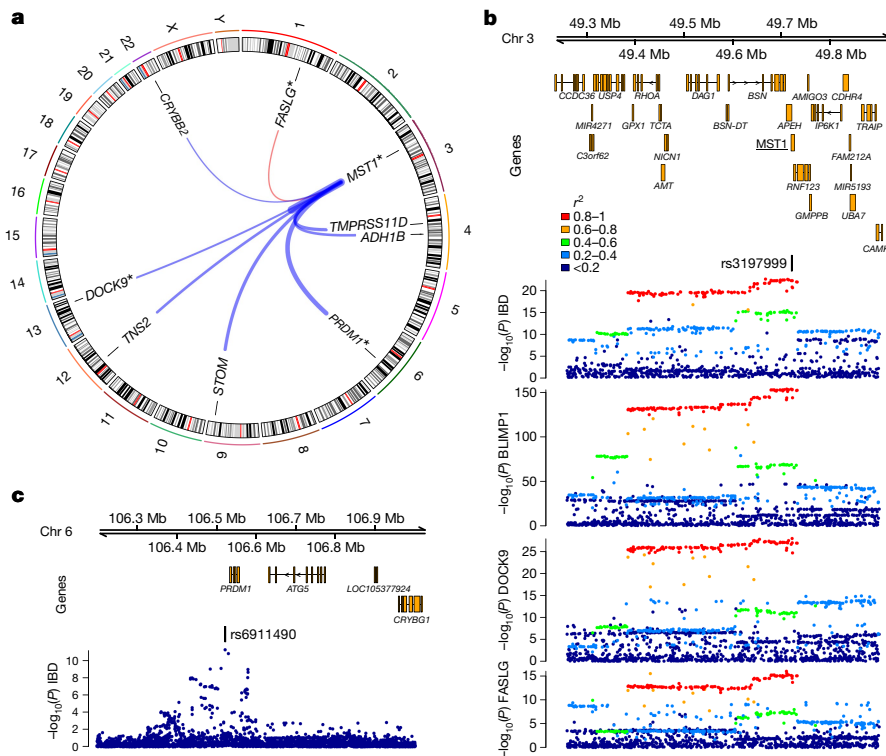


**Fig. 2 | Missense variant rs28929474:T in *SERPINA1* is a *trans* pQTL hotspot.** Outermost numbers indicate chromosomes. Lines link the genomic location of rs28929474 with genes encoding significantly associated proteins. Associations with and without asterisks indicate significance at  $P < 5 \times 10^{-8}$  and  $P < 1.5 \times 10^{-11}$ , respectively. Line thickness is proportional to effect size (red, positive; blue, negative);  $n = 3,301$  participants.

a further six at  $P < 5 \times 10^{-8}$  (Fig. 2). This variant (the ‘Z-allele’) results in defective secretion and intracellular accumulation of  $\alpha_1$ -antitrypsin (A1AT), an anti-protease. Individuals homozygous for the Z allele have a deficiency of circulating A1AT and an increased risk of emphysema, liver cirrhosis and vasculitis. The ‘protease–antiprotease’ hypothesis posits that the pulmonary manifestations of A1AT deficiency result from unchecked protease activity. Our discovery of multiple *trans*-associated proteins at this locus highlights additional pathways that might be relevant to pathogenesis, a hypothesis supported by accumulating data<sup>21</sup>.

GWAS have identified thousands of loci associated with common diseases, but the mechanisms by which most variants influence disease susceptibility are unknown. To identify intermediate links between genotype and disease, we overlapped pQTLs with disease-associated variants from GWAS. Eighty-eight of our sentinel pQTL variants were in high LD ( $r^2 \geq 0.8$ ) with sentinel disease-associated variants (Supplementary Table 14), including 30 with *cis* associations, 54 with *trans*, and 4 with both. As some genetic loci are associated with multiple diseases, these 88 variants represent 253 distinct genotype–disease associations. Overlap of a pQTL and a disease association does not necessarily imply that the same genetic variant underlies both traits, because there may be distinct causal variants for each trait that are in LD. We therefore performed colocalization testing (see Methods). Of 108 locus–disease associations outside the major histocompatibility (MHC) region for which testing was possible, colocalization was highly likely (PP  $> 0.8$ ) for 96 (88.9%), and the most likely explanation (PP  $> 0.5$ ) for 106 (98.1%) (Supplementary Table 14).

*trans* pQTLs that overlap with disease associations can highlight previously unsuspected candidate proteins through which genetic loci may influence disease risk. To help to identify such candidates, we applied the ProGeM framework<sup>22</sup> (Methods, Supplementary Table 12, Extended Data Fig. 7). We show that an inflammatory bowel disease (IBD) risk allele<sup>23</sup> (rs3197999:A, p.Arg703Cys) in *MST1* on chromosome 3, which decreases plasma *MST1* levels<sup>24</sup>, is a *trans* pQTL for eight additional proteins (Supplementary Table 4, Fig. 3). Notably, genes that encode three of these proteins (*PRDM1*, *FASLG* and *DOCK9*)



**Fig. 3 | *trans* pQTL for BLIMP1 at an inflammatory bowel disease (IBD) associated missense variant (rs3197999:A) in *MST1*.**

**a**, rs3197999:A is associated with multiple proteins. Lines link rs3197999 and the genes encoding significantly associated proteins. Line thickness is proportional to effect size of the IBD risk allele (red = positive, blue = negative).  $n = 3,301$  participants. Asterisks indicate genes in

each lie within 500 kb of IBD GWAS loci at which the causal gene is ambiguous<sup>25</sup>. For instance, the IBD-associated variant rs6911490 lies on chromosome 6 in the intergenic region near *PRDM1* (encoding BLIMP1, a master regulator of immune cell differentiation) and *ATG5* (involved in autophagy) (Fig. 3c). Neither fine-mapping nor eQTL colocalization analyses have unequivocally resolved the causal gene at this locus<sup>25</sup>; both *PRDM1* and *ATG5* are plausible candidates. Our data provide support for *PRDM1*.

Anti-neutrophil cytoplasmic antibody-associated vasculitis (AAV) is an autoimmune disease characterized by vascular inflammation and autoantibodies to the neutrophil proteases proteinase-3 (PR3) or myeloperoxidase. GWAS have revealed distinct genetic associations according to antibody specificity<sup>26</sup>, with variants near *PRTN3* (encoding PR3) and at the Z-allele of *SERPINA1* (encoding A1AT, which inhibits PR3) associated specifically with PR3-antibody positive AAV. The SOMAscan assay has two SOMAmers that target PR3; we identified a *cis* pQTL immediately upstream of *PRTN3* for both, and replicated it with the Olink assay (Supplementary Table 4, Fig. 4a, b). Conditional analysis revealed multiple independently associated variants (Supplementary Table 5), one of which (rs7254911) was in high LD with the previously reported<sup>26,27</sup> PR3+ vasculitis-associated variants in the *PRTN3* region (Supplementary Note). We show that the vasculitis risk allele at *PRTN3* is associated with higher plasma levels of PR3 (Supplementary Note Table 4).

For one PR3 SOMAmer, we also found a *trans* pQTL at *SERPINA1*, with the Z-allele being associated with reduced levels of plasma PR3 (Fig. 4a). To understand the SOMAmer-specific nature of this association, we assayed the relative affinity of these SOMAmers for the free and complexed forms of PR3 and A1AT. We found that the SOMAmer showing *cis* and *trans* associations predominantly measured the PR3–A1AT complex rather than free PR3, whereas the SOMAmer with only a *cis* association measured both the free and complexed forms (Extended Data Fig. 8, Supplementary Note). Notably, neither

IBD GWAS loci. **b**, Regional association plots at *MST1*, showing IBD association (top) and *trans* pQTLs for BLIMP1, DOCK9 and FASLG. Colour key indicates  $r^2$  with rs3197999. **c**, Regional association plot of the IBD susceptibility locus at *PRDM1*, which encodes BLIMP1. IBD association data are for European participants from a GWAS meta-analysis<sup>23</sup>.

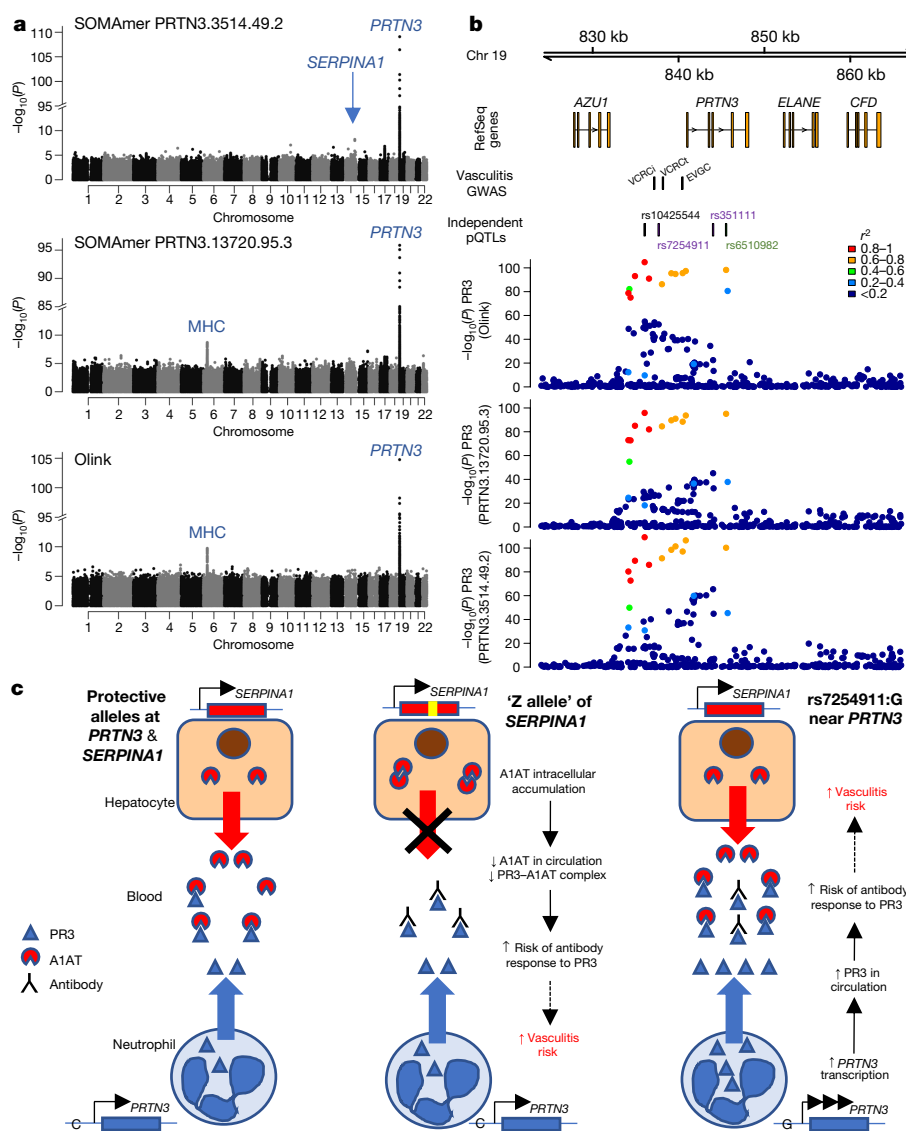
SOMAmer bound free A1AT, demonstrating that the *SERPINA1* pQTL did not reflect non-specific cross-reactivity (Supplementary Note).

These data show that the vasculitis risk allele at *PRTN3* increases total PR3 plasma levels, consistent with its effect on *PRTN3* mRNA abundance in whole blood in GTEx data<sup>28</sup>. The *SERPINA1* Z-allele results in a reduced proportion of PR3 bound to A1AT. We thus demonstrate that altered availability of PR3, conferred by two independent genetic mechanisms, is a key susceptibility factor for breaking immune tolerance to PR3 and the development of PR3+ vasculitis (Fig. 4c).

### Causal evaluation of candidate proteins in disease

Association of plasma protein levels with disease risk does not necessarily imply causation. To help to establish causality, we used Mendelian randomization (MR) analysis<sup>11</sup>, which uses genetic variants as instrumental variables to avoid confounding and reverse causation (Extended Data Fig. 9). If a genetic variant is specifically associated with levels of a protein and is also associated with disease risk, then this provides evidence of the protein's causal role. For example, serum levels of PSP-94 (also known as MSMB) are lower in men who go on to develop prostate cancer<sup>29</sup>, but it is unclear whether this association is correlative or causal. We identified a *cis* pQTL associated with lower PSP-94 plasma levels that overlaps with the prostate cancer susceptibility variant rs10993994<sup>30</sup>, supporting a protective role for PSP-94 in prostate cancer (Supplementary Table 14).

Next, we leveraged multi-variant MR analysis methods to identify causal proteins among multiple plausible candidates, exemplified by the *IL1RL1–IL18R1* locus, which is associated with multiple immune-mediated diseases including atopic dermatitis<sup>31</sup>. We identified four proteins that each had *cis* pQTLs at this locus (Supplementary Table 4), and created a genetic score for each protein (see Methods). Initial 'one-protein-at-a-time' analysis identified associations of the scores for IL18R1 ( $P = 9.3 \times 10^{-72}$ ) and IL1RL1 ( $P = 5.7 \times 10^{-27}$ ) with atopic dermatitis risk (Fig. 5a), and a weak association for IL1RL2 ( $P = 0.013$ ).



**Fig. 4 | Proteinase-3, *SERPINA1* and vasculitis.** **a**, Manhattan plots for plasma PR3 measured with two SOMAmers and the Olink assay. **b**, *PRTN3* regional association plots. Colour key indicates  $r^2$  with sentinel variant rs10425544. ‘Vasculitis GWAS’: previously reported vasculitis-associated variants (see Supplementary Note). EVGC, rs62132295 (from European Vasculitis Genetics Consortium<sup>26</sup>); VCRCi, rs138303849 and VCRCt, rs62132293 (most significant imputed and genotyped variants, respectively, from Vasculitis Clinical Research Consortium<sup>27</sup>). ‘Independent pQTLs’: conditionally independent PR3 pQTL variants

(black lettering shows lead variant for both SOMAmers; purple and green show conditionally independent variants for SOMAmers PRTN3.3514.49.2 and PRTN3.13720.95.3, respectively). **c**, Proposed mechanisms by which *PRTN3* and *SERPINA1* affect PR3 levels and thus vasculitis risk. Left, individuals without either the *PRTN3* or *SERPINA1* vasculitis risk alleles. Middle, *SERPINA1* Z-allele carriers have lower circulating A1AT, resulting in higher free plasma PR3. Right, *cis*-acting variant at the *PRTN3* locus results in higher total plasma PR3. Increases in either free or total PR3 predispose individuals to loss of immune tolerance.

We then mutually adjusted these associations for one another to account for the effects of the variants on multiple proteins. Whereas the association of IL18R1 remained significant ( $P = 1.5 \times 10^{-28}$ ), the association of IL1RL1 ( $P = 0.01$ ) was attenuated. In contrast, the association of IL1RL2 ( $P = 1.1 \times 10^{-69}$ ) became much stronger, suggesting that IL1RL2 and IL18R1 underlie atopic dermatitis risk at this locus.

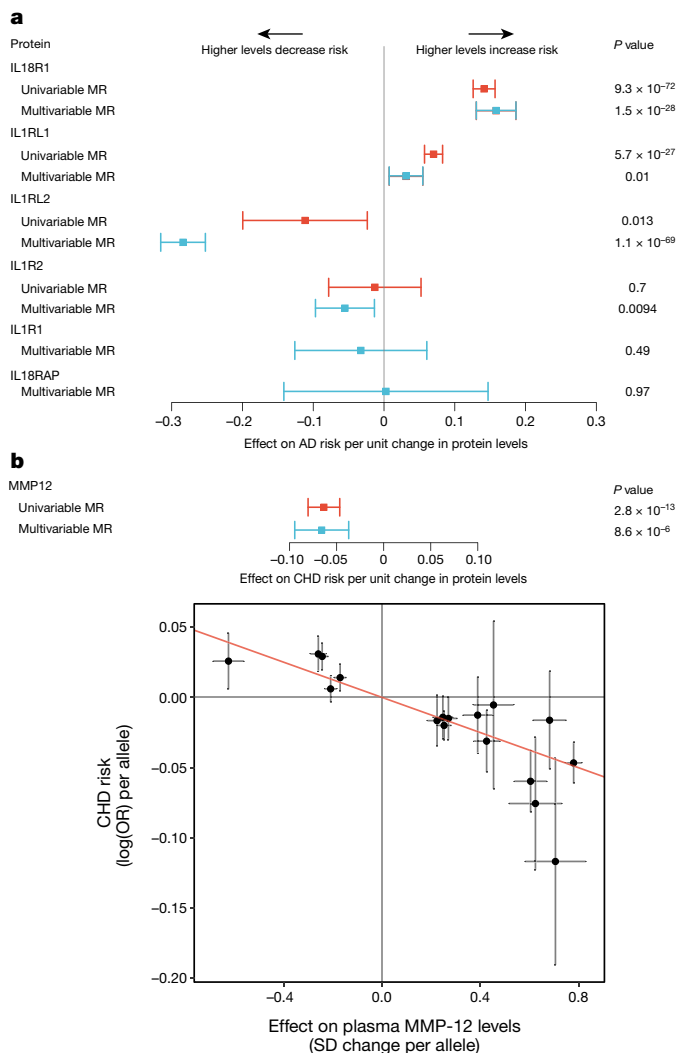
MMP-12 plays a key role in lung tissue damage, and MMP-12 inhibitors are being tested as treatments for chronic obstructive pulmonary disease<sup>32</sup>. We created a multi-allelic genetic score that explains 14% of the variation in plasma MMP-12 levels (see Methods). Observational studies reveal that higher levels of plasma MMP-12 are associated with recurrent cardiovascular events<sup>33</sup>, stimulating interest in the use of MMP-12 inhibitors to treat cardiovascular disease. However, we found that genetic predisposition to higher MMP-12 levels is associated with decreased coronary disease risk ( $P = 2.8 \times 10^{-13}$ ) (Fig. 5b) and decreased large artery atherosclerotic stroke risk<sup>34</sup>. It will be important to understand the discordance between the observational

epidemiology and the genetic risk score, given the therapeutic interest in this target.

### Drug target prioritization

Drugs directed at targets with human genetic support have a greater likelihood of therapeutic success than those directed at unsupported targets<sup>35</sup>. Of the proteins for which we identified a pQTL, 244 (17%) are established drug targets in the Informa Pharmaprojects database (Supplementary Table 15). Thirty-one pQTLs for drug target proteins were highly likely to colocate (PP > 0.8) with a GWAS disease locus, including some that are targets of approved drugs such as tocilizumab (anti-IL6R) and ustekinumab (anti-IL12/23) (Supplementary Table 16a).

To identify additional indications for existing drugs, we investigated disease associations of pQTLs for proteins already targeted by licensed drugs. Our results suggest potential drug repurposing opportunities. For example, we identified a *cis* pQTL for RANK (encoded by



**Fig. 5 | Evaluation of causal role of proteins in disease.**  $n = 3,301$  participants. **a**, MR estimates with 95% confidence intervals (CIs) (instrumental variable analysis) for proteins encoded in the *IL1RL1*–*IL18R1* locus and risk of atopic dermatitis (AD) risk. Univariable MR not possible for *IL1R1* and *IL18RAP* (no significant pQTLs to select as ‘genetic instruments’). **b**, MMP-12 levels and risk of coronary heart disease (CHD). Top, MR estimates with 95% CIs. Bottom, estimated effect sizes (with 95% CIs) on plasma MMP-12 (from linear regression) and CHD risk (from logistic regression) for each variant used in the genetic score.

*TNFRSF11A*) at rs884205, a variant associated with Paget’s disease<sup>36</sup>, which is characterized by excessive bone turnover, deformity and fracture (Supplementary Table 16b). The standard treatment for Paget’s disease is osteoclast inhibition with bisphosphonates, originally developed as anti-osteoporotic drugs. Denosumab, another anti-osteoporosis drug, is a monoclonal antibody targeting RANKL, the ligand for RANK. Our data suggest that denosumab may be an alternative treatment for Paget’s disease when bisphosphonates are contraindicated, a hypothesis supported by clinical case reports<sup>37</sup>.

Next we evaluated targets of drugs currently under development. Drugs targeting GPIBA, a receptor for von Willebrand factor, are in preclinical development as anti-thrombotic agents and in phase 2 trials for thrombotic thrombocytopenic purpura. We found a *cis* pQTL associated with both higher GPIBA abundance and higher platelet count, suggesting a link between GPIBA and platelet count (Supplementary Table 16). Furthermore, we identified a *trans* pQTL for GPIBA at the *SH2B3*–*BRAP* locus, which colocalized with associations with platelet count<sup>10</sup>, myocardial infarction and stroke (Supplementary Table 16b). The risk allele for cardiovascular disease increases both plasma GPIBA

and platelet count, suggesting that GPIBA influences vascular risk via platelets. Collectively, these results support targeting GPIBA in conditions characterized by platelet aggregation such as arterial thrombosis. More generally, our data provide a substrate for generating hypotheses about potential therapeutic targets through linking genetic factors to disease via specific proteins.

## Discussion

This study elucidates the genetic control of the human plasma proteome and uncovers intermediate molecular pathways that connect the genome to disease endpoints. We applied our discoveries to evaluate causal roles for proteins in human diseases using the principle of Mendelian randomization. Proteins provide an ideal paradigm for MR analysis because they are under proximal genetic control. However, application of protein-based MR has been constrained by limited availability of suitable genetic instruments, a bottleneck remedied by our approach. Our study provides a resource for understanding complex traits and an example of the application of novel bioassay technologies to population biobanks.

## Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0175-2>

Received: 28 March 2017; Accepted: 27 April 2018;  
Published online 6 June 2018.

- Albert, F. W. & Kruglyak, L. The role of regulatory variation in complex traits and disease. *Nat. Rev. Genet.* **16**, 197–212 (2015).
- Liu, Y. et al. Quantitative variability of 342 plasma proteins in a human twin population. *Mol. Syst. Biol.* **11**, 786 (2015).
- Suhre, K. et al. Connecting genetic risk to disease end points through the human blood plasma proteome. *Nat. Commun.* **8**, 14357 (2017).
- Yao, C. et al. Genome-wide association study of plasma proteins identifies putatively causal genes, proteins, and pathways for cardiovascular disease. Preprint at <https://www.biorxiv.org/content/early/2017/05/12/136523> (2017).
- de Vries, P. S. et al. Whole-genome sequencing study of serum peptide levels: the Atherosclerosis Risk in Communities study. *Hum. Mol. Genet.* **26**, 3442–3450 (2017).
- Wu, L. et al. Variation and genetic control of protein abundance in humans. *Nature* **499**, 79–82 (2013).
- Battle, A. et al. Impact of regulatory variation from RNA to protein. *Science* **347**, 664–667 (2015).
- Rohloff, J. C. et al. Nucleic acid ligands with protein-like side chains: modified aptamers and their use as diagnostic and therapeutic agents. *Mol. Ther. Nucleic Acids* **3**, e201 (2014).
- Di Angelantonio, E. et al. Efficiency and safety of varying the frequency of whole blood donation (INTERVAL): a randomised trial of 45 000 donors. *Lancet* **390**, 2360–2371 (2017).
- Astle, W. J. et al. The allelic landscape of human blood cell trait variation and links to common complex disease. *Cell* **167**, 1415–1429.e19 (2016).
- Burgess, S., Scott, R. A., Timpson, N. J., Davey Smith, G. & Thompson, S. G. Using published data in Mendelian randomization: a blueprint for efficient identification of causal risk factors. *Eur. J. Epidemiol.* **30**, 543–552 (2015).
- Stranger, B. E. et al. Patterns of *cis* regulatory variation in diverse human populations. *PLoS Genet.* **8**, e1002639 (2012).
- Lundberg, M., Eriksson, A., Tran, B., Assarsson, E. & Fredriksson, S. Homogeneous antibody-based proximity extension assays provide sensitive and specific detection of low-abundant proteins in human blood. *Nucleic Acids Res.* **39**, e102 (2011).
- Walter, K. et al. The UK10K project identifies rare variants in health and disease. *Nature* **526**, 82–90 (2015).
- Westra, H.-J. et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat. Genet.* **45**, 1238–1243 (2013).
- Lappalainen, T. et al. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* **501**, 506–511 (2013).
- Schadt, E. E. et al. Mapping the genetic architecture of gene expression in human liver. *PLoS Biol.* **6**, e107 (2008).
- Zeller, T. et al. Genetics and beyond—the transcriptome of human monocytes and disease susceptibility. *PLoS ONE* **5**, e10693 (2010).
- Nai, A. et al. TMPRSS6 rs855791 modulates hepcidin transcription in vitro and serum hepcidin levels in normal individuals. *Blood* **118**, 4459–4462 (2011).
- Carrasquillo, M. M. et al. Genome-wide screen identifies rs646776 near sortilin as a regulator of progranulin levels in human plasma. *Am. J. Hum. Genet.* **87**, 890–897 (2010).

21. Gooptu, B., Dickens, J. A. & Lomas, D. A. The molecular and cellular pathology of  $\alpha_1$ -antitrypsin deficiency. *Trends Mol. Med.* **20**, 116–127 (2014).
  22. Stacey, D. et al. ProGeM: A framework for the prioritisation of candidate causal genes at molecular quantitative trait loci. <https://doi.org/10.1101/230094> (2017).
  23. Liu, J. Z. et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat. Genet.* **47**, 979–986 (2015).
  24. Di Narzo, A. F. et al. High-throughput characterization of blood serum proteomics of IBD patients with respect to aging and genetic factors. *PLoS Genet.* **13**, e1006565 (2017).
  25. Huang, H. et al. Fine-mapping inflammatory bowel disease loci to single-variant resolution. *Nature* **547**, 173–178 (2017).
  26. Lyons, P. A. et al. Genetically distinct subsets within ANCA-associated vasculitis. *N. Engl. J. Med.* **367**, 214–223 (2012).
  27. Merkel, P. A. et al. Identification of functional and expression polymorphisms associated with risk for anti-neutrophil cytoplasmic autoantibody-associated vasculitis. *Arthritis Rheumatol.* **69**, 1054–1066 (2017).
  28. Battle, A., Brown, C. D., Engelhardt, B. E. & Montgomery, S. B. Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).
  29. Grönberg, H. et al. Prostate cancer screening in men aged 50–69 years (STHLM3): a prospective population-based diagnostic study. *Lancet Oncol.* **16**, 1667–1676 (2015).
  30. Eeles, R. A. et al. Multiple newly identified loci associated with prostate cancer susceptibility. *Nat. Genet.* **40**, 316–321 (2008).
  31. Paternoster, L. et al. Multi-ancestry genome-wide association study of 21,000 cases and 95,000 controls identifies new risk loci for atopic dermatitis. *Nat. Genet.* **47**, 1449–1456 (2015).
  32. Dahl, R. et al. Effects of an oral MMP-9 and -12 inhibitor, AZD1236, on biomarkers in moderate/severe COPD: a randomised controlled trial. *Pulm. Pharmacol. Ther.* **25**, 169–177 (2012).
  33. Ganz, P. et al. Development and validation of a protein-based risk score for cardiovascular outcomes among patients with stable coronary heart disease. *J. Am. Med. Assoc.* **315**, 2532–2541 (2016).
  34. Traylor, M. et al. A novel MMP12 locus is associated with large artery atherosclerotic stroke using a genome-wide age-at-onset informed approach. *PLoS Genet.* **10**, e1004469 (2014).
  35. Nelson, M. R. et al. The support of human genetic evidence for approved drug indications. *Nat. Genet.* **47**, 856–860 (2015).
  36. Albagha, O. M. E. et al. Genome-wide association study identifies variants at CSF1, OPTN and TNFRSF11A as genetic risk factors for Paget's disease of bone. *Nat. Genet.* **42**, 520–524 (2010).
  37. Schwarz, P., Rasmussen, A. Q., Kvist, T. M., Andersen, U. B. & Jørgensen, N. R. Paget's disease of the bone after treatment with Denosumab: a case report. *Bone* **50**, 1023–1025 (2012).
- Acknowledgements** A. Day-Williams, J. McElwee, D. Diogo, W. Astle, E. Di Angelantonio, E. Birney, A. Richard, J. Mason and M. Inouye commented on the manuscript, and M. Sharp helped with mapping drug indications to GWAS traits. We thank INTERVAL study participants; staff at recruiting NHSBT blood donation centres; and the INTERVAL Study Co-ordination team, Operations Team (led by R. Houghton and C. Moore) and Data Management Team (led by M. Walker). Funding sources are listed in the Supplementary Information.
- Reviewer information** *Nature* thanks T. Lappalainen, M. McCarthy and the other anonymous reviewer(s) for their contribution to the peer review of this work.
- Author contributions** Conceptualization and experimental design: J.D., A.S.B., B.B.S., H.R., R.M.P.; methodology: B.B.S., A.S.B., J.C.M., J.E.P., H.R., S.B.; conducted experimental work: N.J., S.K.W., E.S.Z.; analysis: B.B.S., J.C.M., J.E.P., D.S., J.B., J.R.S., T.J., E.P., P.S., C.O.-W., M.A.K., S.K.W., A.C., N.B., S.L.S.; contributed reagents, materials, protocols or analysis tools: N.J., S.K.W., E.S.Z., J.B., M.A.K., J.R.S., B.P.P.; supervision: A.S.B., H.R., J.D., R.M.P., C.S.F., D.S.P., A.M.W.; writing: A.S.B., J.E.P., B.B.S., J.C.M., H.R., J.D., J.A.T., N.S., K.S.; creation of the INTERVAL BioResource: J.R.B., D.J.R., W.H.O., N.W.M., J.D.; funding: N.W.M., J.R.B., D.J.R., W.H.O., H.R., R.M.P., J.D.; all authors critically reviewed the manuscript.
- Competing interests** The authors declare the following competing interests: A.C., CSF-Merck employee; N.J., S.K.W., SomaLogic Inc employees and stakeholders; E.S.Z., SomaLogic Inc employee; J.C.M., R.M.P., Merck employees during this study, now Celgene employees; H.R., Merck employee during this study; J.E.P., travel and accommodation expenses and hospitality from Olink to speak at Olink-sponsored academic meetings; A.S.B., grants from Merck, Pfizer, Novartis, Biogen and Bioverativ and personal fees from Novartis; J.D., sits on the Novartis Cardiovascular and Metabolic Advisory Board, had grant support from Novartis.
- Additional information**
- Extended data** is available for this paper at <https://doi.org/10.1038/s41586-018-0175-2>.
- Supplementary information** is available for this paper at <https://doi.org/10.1038/s41586-018-0175-2>.
- Reprints and permissions information** is available at <http://www.nature.com/reprints>.
- Correspondence and requests for materials** should be addressed to J.D. or A.S.B.
- Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## METHODS

**Study participants.** The INTERVAL study comprises about 50,000 participants nested within a randomized trial of varying blood donation intervals<sup>9</sup>. Between mid-2012 and mid-2014, blood donors aged 18 years and older were recruited at 25 centres of England's National Health Service Blood and Transplant (NHSBT). All participants gave informed consent before joining the study and the National Research Ethics Service approved this study (11/EE/0538). Participants completed an online questionnaire including questions about demographic characteristics (for example, age, sex, ethnicity), anthropometry (height, weight), lifestyle (for example, alcohol and tobacco consumption) and diet. Participants were generally in good health because blood donation criteria exclude people with a history of major diseases (such as myocardial infarction, stroke, cancer, HIV, and hepatitis B or C) and those who have had recent illness or infection. For SomaLogic assays, we randomly selected two non-overlapping subcohorts of 2,731 and 831 participants from INTERVAL. After genetic quality control, 3,301 participants (2,481 and 820 in the two subcohorts) remained for analysis (Supplementary Table 17). No statistical methods were used to determine sample size. The experiments were not randomized. Laboratory staff conducting proteomic assays were blinded to the genotypes of participants.

**Plasma sample preparation.** Sample collection procedures for INTERVAL have been described previously<sup>38</sup>. In brief, blood samples for research purposes were collected in 6-ml EDTA tubes using standard venepuncture protocols. The tubes were inverted three times and transferred at ambient temperature to UK Biocentre (Stockport, UK) for processing. Plasma was extracted into two 0.8-ml plasma aliquots by centrifugation and subsequently stored at  $-80^{\circ}\text{C}$  before use.

**Protein measurements.** We used a multiplexed, aptamer-based approach (SOMAscan assay) to measure the relative concentrations of 3,622 plasma proteins or protein complexes assayed using 4,034 modified aptamers ('SOMAmers', hereafter referred to as SOMAmers; Supplementary Table 18). The assay extends the lower limit of detectable protein abundance afforded by conventional approaches (for example, immunoassays), measuring both extracellular and intracellular proteins (including soluble domains of membrane-associated proteins), with a bias towards proteins likely to be found in the human secretome<sup>8,39</sup> (Extended Data Fig. 10a). The proteins cover a wide range of molecular functions (Extended Data Fig. 10b). The selection of proteins on the platform reflects both the availability of purified protein targets and a focus on proteins suspected to be involved in the pathophysiology of human disease.

Aliquots of 150  $\mu\text{l}$  of plasma were sent on dry ice to SomaLogic Inc. (Boulder, Colorado, US) for protein measurement. Assay details have been previously described<sup>39,40</sup> and a technical white paper with further information can be found at the manufacturer's website ([http://somallogic.com/wp-content/uploads/2017/06/SSM-002-Technical-White-Paper\\_010916\\_LSM1.pdf](http://somallogic.com/wp-content/uploads/2017/06/SSM-002-Technical-White-Paper_010916_LSM1.pdf)). In brief, modified single-stranded DNA SOMAmers are used to bind to specific protein targets that are then quantified using a DNA microarray. Protein concentrations are quantified as relative fluorescent units.

Quality control (QC) was performed at the sample and SOMAmer levels using control aptamers and calibrator samples. At the sample level, hybridization controls on the microarray were used to correct for systematic variability in hybridization, while the median signal over all features assigned to one of three dilution sets (40%, 1% and 0.005%) was used to correct for within-run technical variability. The resulting hybridization scale factors and median scale factors were used to normalize data across samples within a run. The acceptance criteria for these values are between 0.4 and 2.5 based on historical runs. SOMAmer-level QC made use of replicate calibrator samples using the same study matrix (plasma) to correct for between-run variability. The acceptance criterion for each SOMAmer was that the calibration scale factor be less than 0.4 from the median for each of the plates run. In addition, at the plate level, the acceptance criteria were that the median of the calibration scale factors be between 0.8 and 1.2, and that 95% of individual SOMAmers be less than 0.4 from the median within the plate.

In addition to QC processes routinely conducted by SomaLogic, we measured protein levels of 30 and 10 pooled plasma samples randomly distributed across plates for subcohort 1 and subcohort 2, respectively. Laboratory technicians were blinded to the presence of pooled samples. This approach enabled estimation of the reproducibility of the protein assays. We calculated the coefficient of variation (CV) for each SOMAmer within each subcohort by dividing the standard deviation by the mean of the pooled plasma sample protein read-outs. In addition to passing SomaLogic QC processes, we required SOMAmers to have a CV  $\leq 20\%$  in both subcohorts. Eight non-human protein targets were also excluded, leaving 3,283 SOMAmers (mapping to 2,994 unique proteins or protein complexes) for inclusion in the GWAS.

Protein mapping to UniProt identifiers and gene names was provided by SomaLogic. Mapping to Ensembl gene IDs and genomic positions was performed using Ensembl Variant Effect Predictor v83 (VEP)<sup>41</sup>. Protein subcellular locations were determined by exporting the subcellular location annotations from UniProt<sup>42</sup>.

If the term 'membrane' was included in the descriptor, the protein was considered to be a membrane protein, whereas if the term 'secreted' (but not 'membrane') was included in the descriptor, the protein was considered to be a secreted protein. Proteins not annotated as either membrane or secreted proteins were classified (by inference) as intracellular proteins. Proteins were mapped to molecular functions using gene ontology annotations<sup>43</sup> from UniProt.

**Non-genetic associations of proteins.** To provide confidence in the reproducibility of the protein assays, we attempted to replicate the associations with age or sex of 45 proteins previously reported by Ngo et al. and 40 reported by Menni et al.<sup>44,45</sup>. We used Bonferroni-corrected  $P$  value thresholds of  $P = 1.1 \times 10^{-3}$  (0.05/45) and  $P = 1.2 \times 10^{-3}$  (0.05/40), respectively. Relative protein abundances were rank-inverse normalized within each subcohort and linear regression was performed using age, sex, body mass index, natural log of estimated glomerular filtration rate (eGFR) and subcohort as independent variables.

**Genotyping and imputation.** The genotyping protocol and QC for the INTERVAL samples ( $n \approx 50,000$ ) have been described previously in detail<sup>10</sup>. DNA extracted from buffy coat was used to assay approximately 830,000 variants on the Affymetrix Axiom UK Biobank genotyping array at Affymetrix (Santa Clara, California, US). Genotyping was performed in multiple batches of approximately 4,800 samples each. Sample QC was performed including exclusions for sex mismatches, low call rates, duplicate samples, extreme heterozygosity and non-European descent. Relatedness was removed by excluding one participant from each pair of close (first- or second-degree) relatives, defined as  $\hat{\pi} > 0.187$ . Identity-by-descent was estimated using a subset of variants with a call rate  $>99\%$  and MAF  $> 5\%$  in the merged data set of both subcohorts, pruned for linkage disequilibrium (LD) using PLINK v1.9<sup>46</sup>. Numbers of participants excluded at each stage of the genetic QC are summarized in Extended Data Fig. 1. Multi-dimensional scaling was performed using PLINK v1.9 to create components to account for ancestry in genetic analyses.

Prior to imputation, additional variant filtering steps were performed to establish a high-quality imputation scaffold. In summary, 654,966 high-quality variants (autosomal, non-monomorphic, bi-allelic variants with Hardy-Weinberg Equilibrium (HWE)  $P > 5 \times 10^{-6}$ , with a call rate of  $>99\%$  across the INTERVAL genotyping batches in which a variant passed QC, and a global call rate of  $>75\%$  across all INTERVAL genotyping batches) were used for imputation. Variants were phased using SHAPEIT3 and imputed using a combined 1000 Genomes Phase 3-UK10K reference panel. Imputation was performed via the Sanger Imputation Server (<https://imputation.sanger.ac.uk>) and resulted in 87,696,888 imputed variants.

Prior to genetic association testing, variants were filtered in each subcohort separately using the following exclusion criteria: (1) imputation quality (INFO) score  $< 0.7$ ; (2) minor allele count  $< 8$ ; (3) HWE  $P < 5 \times 10^{-6}$ . In the small number of cases in which imputed variants had the same genomic position (GRCh37) and alleles, the variant with the lowest INFO score was removed. 10,572,788 variants passing all filters in both subcohorts were taken forward for analysis (Extended Data Fig. 1).

**Genome-wide association study.** Within each subcohort, relative protein abundances were first natural log-transformed. Log-transformed protein levels were then adjusted in a linear regression for age, sex, duration between blood draw and processing (binary,  $\leq 1$  day/ $> 1$  day) and the first three principal components of ancestry from multi-dimensional scaling. The protein residuals from this linear regression were then rank-inverse normalized and used as phenotypes for association testing. Simple linear regression using an additive genetic model was used to test genetic associations. Association tests were carried out on allelic dosages to account for imputation uncertainty ('-method expected' option) using SNPTTEST v2.5.2<sup>47</sup>.

**Meta-analysis and statistical significance.** Association results from the two subcohorts were combined via fixed-effects inverse-variance meta-analysis combining the betas and standard errors using METAL<sup>48</sup>. Genetic associations were considered to be genome-wide significant based on a conservative strategy requiring associations to have (i) a meta-analysis  $P$  value  $< 1.5 \times 10^{-11}$  (genome-wide threshold of  $P = 5 \times 10^{-8}$  Bonferroni-corrected for 3,283 aptamers tested), (ii) at least nominal significance ( $P < 0.05$ ) in both subcohorts, and (iii) consistent direction of effect across subcohorts. We did not observe significant genomic inflation (mean inflation factor was 1.0, standard deviation = 0.01) (Extended Data Fig. 3d).

**Refinement of significant regions.** To identify distinct non-overlapping regions associated with a given SOMAmer, we first defined a 1-Mb region around each significant variant for that SOMAmer. Starting with the region containing the variant with the smallest  $P$  value, any overlapping regions were then merged and this process was repeated until no more overlapping 1-Mb regions remained. The variant with the lowest  $P$  value for each region was assigned as the 'regional sentinel variant'. Owing to the complexity of the MHC region, we treated the extended MHC region (chr6:25.5–34.0Mb) as one region. To identify whether a region was associated with multiple SOMAmers, we used an LD-based clumping approach. Regional sentinel variants in high LD ( $r^2 \geq 0.8$ ) with each other were combined together into a single region.

**Conditional analyses.** To identify conditionally significant associations, we performed approximate genome-wide stepwise conditional analysis using GCTA v1.25.2<sup>49</sup> using the 'cojo-slc' option. We used the same conservative significance threshold of  $P = 1.5 \times 10^{-11}$  as for the univariable analysis. As inputs for GCTA, we used the summary statistics (that is, betas and standard errors) from the meta-analysis. Correlation between variants was estimated using the 'hard-called' genotypes (where a genotype was called if it had a posterior probability of  $>0.9$  following imputation or set to missing otherwise) in the merged genetic data set, and only variants also passing the univariable genome-wide threshold ( $P < 1.5 \times 10^{-11}$ ) were considered for stepwise selection. As the conditional analyses use different data inputs to the univariable analysis (that is, summarized rather than individual-level data), there were some instances where the conditional analysis failed to include in the stepwise selection sentinel variants that were only just statistically significant in the univariable analysis. In these instances ( $n = 28$ ), we re-conducted the joint model estimation without stepwise selection in GCTA, using the variants identified by the conditional analysis in addition to the regional sentinel variant. We report and highlight these cases in Supplementary Table 5.

**Replication of previous pQTLs.** We attempted to identify all previously reported pQTLs from GWAS and to assess whether they replicated in our study. We used the NCBI Entrez programming utility in R (rentrez) to perform a literature search for pQTL studies published from 2008 onwards. We searched for the following terms: 'pQTL', 'pQTLs', and 'protein quantitative trait locus'. We supplemented this search by filtering out GWAS associations from the NHGRI-EBI GWAS Catalog v1.0.1<sup>50</sup> (<https://www.ebi.ac.uk/gwas/>, downloaded November 2017), which has all phenotypes mapped to the Experimental Factor Ontology (EFO)<sup>51</sup>, by restricting to those with EFO annotations relevant to protein biomarkers (for example, 'protein measurement', EFO\_0004747). Studies identified through both approaches were manually filtered to include only studies that profiled plasma or serum samples and to exclude studies not assessing proteins. We recorded basic summary information for each study including the assay used, sample size and number of proteins with pQTLs (Supplementary Table 19). To reduce the impact of ethnic differences in allele frequencies on replication rate estimates, we filtered studies to include only associations reported in European-ancestry populations. We then manually extracted summary data on all reported associations from the manuscript or the supplementary material. This included rsID, protein UniProt ID,  $P$  values, and whether the association was *cis* or *trans* (Supplementary Table 20).

To assess replication we first identified the set of unique UniProt IDs that were also assayed on the SOMAscan panel. For previous studies that used SomaLogic technology, we refined this match to the specific aptamer used. We then clumped associations into distinct loci using the same method that we applied to our pQTLs (see 'Refinement of significant regions'). For each locus, we asked whether the sentinel SNP or a proxy ( $r^2 > 0.6$ ) was associated with the same protein (or aptamer) in our study at a defined significance threshold. For our primary assessment, we used a  $P$  value threshold of  $10^{-4}$  (Supplementary Table 21). We also performed sensitivity analyses to explore factors that influence replication rate (Supplementary Note).

**Replication study using Olink assay.** To test replication of 163 pQTLs for 116 proteins, we performed protein measurements using an alternative assay, that is, a proximity extension assay method (Olink Bioscience, Uppsala, Sweden)<sup>13</sup> in an additional subcohort of 4,998 INTERVAL participants. Proteins were measured using three 92-protein 'panels' – 'inflammatory', 'cvd2' and 'cvd3' (10 proteins were assayed on more than 1 panel). 4,902, 4,947 and 4,987 samples passed quality control for the 'inflammatory', 'cvd2' and 'cvd3' panels, respectively, of which 712, 715 and 721 samples were from individuals included in our primary pQTL analysis using the SOMAscan assay. Normalized protein levels ('NPX') were regressed on age, sex, plate, time from blood draw to processing (in days), and season (categorical: 'Spring', 'Summer', 'Autumn', 'Winter'). The residuals were then rank-inverse normalized. Genotype data was processed as described earlier. Linear regression of the rank-inversed normalized residuals on genotype was carried out in SNPTEST with the first three components of multi-dimensional scaling as covariates to adjust for ancestry. pQTLs were considered to have replicated if they met a  $P$  value threshold Bonferroni-corrected for the number of tests ( $P < 3.1 \times 10^{-4}$ ; 0.05/163) and had a directionally concordant beta estimate with the SOMAscan estimate.

**Candidate gene annotation.** We defined a pQTL as *cis* when the most significantly associated variant in the region was located within 1 Mb of the TSS of the gene(s) encoding the protein. pQTLs lying outside of this region were defined as *trans*. When considering the distance of the lead *cis*-associated variant from the relevant TSS, only proteins that mapped to single genes on the primary assembly in Ensembl v83 were considered.

For *trans* pQTLs, we sought to prioritize candidate genes in the region that might underpin the genotype–protein association. We applied the ProGeM framework<sup>22</sup>, which leverages a combination of databases of molecular pathways, protein–protein interaction networks, and variant annotation, as well as functional genomic data including eQTL and chromosome conformation capture. In addition to reporting the nearest gene to the sentinel variant, ProGeM employs complementary 'bottom

up' and 'top down' approaches, starting from the variant and protein respectively. For the 'bottom up' approach, the sentinel variant and corresponding proxies ( $r^2 > 0.8$ ) for each *trans* pQTL were first annotated using Ensembl VEP v83 (using the 'pick' option) to determine whether variants were (1) protein-altering coding variants; (2) synonymous coding or 5'/3' untranslated region (UTR); (3) intronic or up/downstream; or (4) intergenic. Second, we queried all sentinel variants and proxies against significant *cis* eQTL variants (defined by beta distribution-adjusted empirical  $P$  values using an FDR threshold of 0.05, see <http://www.gtexportal.org/home/documentationPage> for details) in any cell type or tissue from the Genotype-Tissue Expression (GTEx) project v6<sup>28</sup> (<http://www.gtexportal.org/home/datasets>). Third, we also queried promoter capture Hi-C data in 17 human primary haematopoietic cell types<sup>52</sup> to identify contacts (with a CHiCAGO score  $>5$  in at least one cell type) involving chromosomal regions containing a sentinel variant. We considered gene promoters annotated on either fragment (that is, the fragment containing the sentinel variant or the other corresponding fragment) as potential candidate genes. Using these three sources of information, we generated a list of candidate genes for the *trans* pQTLs. A gene was considered a candidate if it fulfilled at least one of the following criteria: (1) it was proximal (intragenic or  $\pm 5$  kb from the gene) or nearest to the sentinel variant; (2) it contained a sentinel or proxy variant ( $r^2 > 0.8$ ) that was protein-altering; (3) it had a significant *cis* eQTL in at least one GTEx tissue overlapping with a sentinel pQTL variant (or proxy); or (4) it was regulated by a promoter annotated on either fragment of a chromosomal contact<sup>52</sup> involving a sentinel variant.

For the 'top down' approach, we first identified all genes with a TSS located within the corresponding pQTL region using the GenomicRanges Bioconductor package<sup>53</sup> with annotation from a GRCh37 GTF file from Ensembl ([ftp://ftp.ensembl.org/pub/grch37/update/gtf/homo\\_sapiens/](ftp://ftp.ensembl.org/pub/grch37/update/gtf/homo_sapiens/); file: 'Homo\_sapiens.GRCh37.82.gtf.gz', downloaded June 2016). We then identified any local genes that had previously been linked with the corresponding *trans*-associated protein(s) according to the following open source databases: (1) the Online Mendelian Inheritance in Man (OMIM) catalogue<sup>54</sup> (<http://www.omim.org/>); (2) the Kyoto Encyclopedia of Genes and Genomes (KEGG)<sup>55</sup> (<http://www.genome.jp/kegg/>); and (3) STRINGdb<sup>56</sup> (<http://string-db.org/>; v10.0). We accessed OMIM data via HumanMine web tool<sup>57</sup> (<http://www.humanmine.org/>; accessed June 2016), whereby we extracted all OMIM IDs for (i) our *trans*-affected proteins and (ii) genes local ( $\pm 500$  kb) to the corresponding *trans*-acting variant. We extracted all human KEGG pathway IDs using the KEGGREST Bioconductor package (<https://bioconductor.org/packages/release/bioc/html/KEGGREST.html>). In cases where a *trans*-associated protein shared either an OMIM ID or a KEGG pathway ID with a gene local to the corresponding *trans*-acting variant, we took this as evidence of a potential functional involvement of that gene. We interrogated protein–protein interaction data by accessing STRINGdb data using the STRINGdb Bioconductor package<sup>58</sup>, whereby we extracted all pairwise interaction scores for each *trans*-affected protein and all proteins with genes local to the corresponding *trans*-acting variants. We took the default interaction score of 400 as evidence of an interaction between the proteins, therefore indicating a possible functional involvement for the local gene. In addition to using data from open source databases in our top down approach, we also adopted a 'guilt-by-association' (GbA) approach using the same plasma proteomic data used to identify our pQTLs. We first generated a matrix containing all possible pairwise Pearson's correlation coefficients between our 3,283 SOMAmers. We then extracted the coefficients relating to our *trans*-associated proteins and any proteins encoded by genes local to their corresponding *trans*-acting variants (where available). Where the correlation coefficient was  $\geq 0.5$  we prioritized the relevant local genes as being potential mediators of the *trans* association(s) at that locus.

We report the potential candidate genes for our *trans* pQTLs from both the 'bottom up' and 'top down' approaches, highlighting cases where the same gene was highlighted by both approaches.

**Functional annotation of pQTLs.** Functional annotation of variants was performed using Ensembl VEP v83 using the 'pick' option. We tested the enrichment of significant pQTL variants for certain functional classes by comparing to permuted sets of variants showing no significant association with any protein ( $P > 0.0001$  for all proteins tested). First, the regional sentinel variants were LD-pruned at  $r^2$  of 0.1. Each time the sentinel variants were LD-pruned, one of the pairs of correlated variants was removed at random and for each set of LD-pruned sentinel variants, 100 equally sized sets of null permuted variants were sampled matching for MAF (bins of 5%), distance to TSS (bins of 0–0.5 kb, 0.5–2 kb, 2–5 kb, 5–10 kb, 10–20 kb, 20–100 kb and  $>100$  kb in each direction) and LD ( $\pm$  half the number of variants in LD with the sentinel variant at  $r^2$  of 0.8). This procedure was repeated 100 times resulting in 10,000 permuted sets of variants. An empirical  $P$  value was calculated as the proportion of permuted variant sets where the proportion that is classified as a particular functional group exceeded that of the test set of sentinel pQTL variants, and we used a significance threshold of  $P = 0.005$  (0.05/10 functional classes tested).

**Evidence against aptamer-binding effects at *cis* pQTLs.** All protein assays that rely on binding (for example, of antibodies or SOMAmers) are susceptible to the possibility of binding-affinity effects, where protein-altering variants (PAVs) (or their proxies in LD) are associated with protein measurements owing to differential binding rather than differences in protein abundance. To account for this potential effect, we performed conditional analysis at all *cis* pQTLs where the sentinel variant was in LD ( $r^2 \geq 0.1$  and  $r^2 \leq 0.9$ ) with a PAV in the gene(s) encoding the associated protein. First, variants were annotated with Ensembl VEP v83 using the 'per-gene' option. Variant annotations were considered protein-altering if they were annotated as coding sequence variant, frameshift variant, in-frame deletion, in-frame insertion, missense variant, protein altering variant, splice acceptor variant, splice donor variant, splice region variant, start lost, stop gained, or stop lost. To avoid multi-collinearity, PAVs were LD-pruned ( $r^2 > 0.9$ ) using PLINK v1.9 before including them as covariates in the conditional analysis on the meta-analysis summary statistics using GCTA v1.25.2. Coverage of known common (MAF  $> 5\%$ ) PAVs in our data was checked by comparison with exome sequences from ~60,000 individuals in the Exome Aggregation Consortium (ExAC (<http://exac.broadinstitute.org>), downloaded June 2016)<sup>59</sup>.

**Testing for regulatory and functional enrichment.** We tested whether our pQTLs were enriched for functional and regulatory characteristics using GARFIELD v1.2.0<sup>60</sup>. GARFIELD is a non-parametric permutation-based enrichment method that compares input variants to permuted sets matched for number of proxies ( $r^2 \geq 0.8$ ), MAF and distance to the closest TSS. It first applies 'greedy pruning' ( $r^2 < 0.1$ ) within a 1-Mb region of the most significant variant. GARFIELD annotates variants with more than a thousand features, drawn predominantly from the GENCODE, ENCODE and ROADMAP projects, which includes genic annotations, histone modifications, chromatin states and other regulatory features across a wide range of tissues and cell types.

The enrichment analysis was run using all variants that passed our Bonferroni-adjusted significance threshold ( $P < 1.5 \times 10^{-11}$ ) for association with any protein. For each of the matching criteria (MAF, distance to TSS, number of LD proxies), we used five bins. In total we tested 25 combinations of features (classified as transcription factor binding sites, FAIRE-seq, chromatin states, histone modifications, footprints, hotspots, or peaks) with up to 190 cell types from 57 tissues, leading to 998 tests. Hence, we considered enrichment with  $P < 5 \times 10^{-5}$  (0.05/998) to be statistically significant.

**Disease annotation.** To identify diseases with which our pQTLs have been associated, we queried our sentinel variants and their strong proxies ( $r^2 \geq 0.8$ ) against publicly available disease GWAS data using PhenoScanner<sup>61</sup>. A list of data sets queried is available at <http://www.phenoscaner.medschl.cam.ac.uk/information.html>. For disease GWAS, results were filtered to  $P < 5 \times 10^{-8}$  and then manually curated to retain only the entry with the strongest evidence for association (that is, smallest  $P$  value) per disease. Non-disease phenotypes such as anthropometric traits, intermediate biomarkers and lipids were excluded manually.

***cis* eQTL overlap and enrichment of *cis* pQTLs for *cis* eQTLs.** For each regional sentinel *cis* pQTL variant, its strong proxies ( $r^2 \geq 0.8$ ) were queried against publicly available eQTL association data using PhenoScanner. *cis* eQTL results were filtered to retain only variants with  $P < 1.5 \times 10^{-11}$ . Only *cis* eQTLs for the same gene as the *cis* pQTL protein were retained. We tested whether *cis* pQTLs were significantly enriched for eQTLs for the corresponding gene compared to null sets of variants appropriately matched for MAF and distance to nearest TSS. For this analysis, we restricted eQTL data to GTEx project v6, since this project provided complete summary statistics across a wide range of tissues and cell-types, in contrast to many other studies which only report  $P$  values below some significance level. GTEx results were filtered to contain only variants lying in *cis* (that is, within 1 Mb) of genes that encode proteins analysed in our study and only variants in both data sets were used.

For the enrichment analysis, the *cis* pQTL sentinel variants were first LD-pruned ( $r^2 < 0.1$ ) and the proportion of sentinel *cis* pQTL variants that are also eQTLs at our pQTL significance threshold ( $P < 1.5 \times 10^{-11}$ ), conventional genome-wide significance ( $P < 5 \times 10^{-8}$ ) or a nominal  $P$  value threshold ( $P < 1 \times 10^{-5}$ ) for the same protein or gene was compared to a permuted set of variants that were not pQTLs ( $P > 0.0001$  for all proteins). We generated 10,000 permuted sets of null variants for each significance threshold matched for MAF, distance to TSS and LD (as described for functional annotation enrichment in 'Functional annotation of pQTLs'). An empirical  $P$  value was calculated as the proportion of permuted variant sets where the proportion that are also *cis* eQTLs exceeded that of the test set of sentinel *cis* pQTL variants.

At a stringent eQTL significance threshold ( $P < 1.5 \times 10^{-11}$ ), we found significant enrichment of *cis* pQTLs for eQTLs ( $P < 0.0001$ ) (Supplementary Table 11) with 19.5% overlap observed compared to a mean overlap of 1.8% in the null sets. Results were similar in sensitivity analyses using the standard genome-wide or nominal significance thresholds as well as when using only the sentinel variants at *cis* pQTLs that were robust to adjusting for PAVs (Supplementary Table 7),

suggesting our results are robust to the choice of threshold and potential differential binding effects.

**Colocalization analysis.** Colocalization testing was performed using the coloc package<sup>62</sup>. For testing colocalization of pQTLs and disease associations, colocalization testing was necessarily limited to disease traits for which full GWAS summary statistics had been made available. We obtained GWAS summary statistics through PhenoScanner. For testing colocalization of pQTLs with eQTLs, we used publicly available summary statistics for expression traits from GTEx<sup>28</sup>. We used the default priors. Regions for testing were determined by dividing the genome into 0.1-cM chunks using recombination data. Evidence for colocalization was assessed using the posterior probability (PP) for hypothesis 4 (that there is an association for both traits and they are driven by the same causal variant(s)). Associations with PP4  $> 0.5$  were deemed likely to colocalize as this gives hypothesis 4 the highest likelihood of being correct, while PP4  $> 0.8$  was deemed to be 'highly likely to colocalize'.

**Selection of genetic instruments for Mendelian randomization.** In MR, genetic variants are used as 'instrumental variables' (IVs) for assessing the causal effect of the exposure (here a plasma protein) on the outcome (here a disease)<sup>11,63</sup> (Extended Data Fig. 9).

**Proteins in the *IL1RL1-IL18R1* locus and atopic dermatitis.** To identify the likely causal proteins that underpin the previous genetic association of the *IL1RL1-IL18R1* locus (chr11:102.5–103.5Mb) with atopic dermatitis (AD)<sup>31</sup>, we used the following approach. For each protein encoded by a gene in the *IL1RL1-IL18R1* locus, we took genetic variants that had a *cis* association at  $P < 1 \times 10^{-4}$  and 'LD-pruned' them at  $r^2 < 0.1$  to leave largely independent variants. We then used these genetic variants to construct a genetic score for each protein. Formally, we used these variants as instrumental variables for their respective proteins in univariable MR. For multivariable MR, association estimates for all proteins in the locus were extracted for all instruments. We used PhenoScanner to obtain association statistics for the selected variants in the European-ancestry population of a recent large-scale GWAS meta-analysis of AD<sup>31</sup>. Where the relevant variant was not available, the strongest proxy with  $r^2 \geq 0.8$  was used.

**MMP-12 and coronary heart disease (CHD).** To test whether plasma MMP-12 levels have a causal effect on risk of CHD, we selected genetic variants in the *MMP12* gene region to use as instrumental variables. We constructed a genetic score comprising 17 variants that had a *cis* association with MMP-12 levels at  $P < 5 \times 10^{-8}$  and that were not highly correlated with one another ( $r^2 < 0.2$ ). To perform multivariable MR, we used association estimates for these variants with other MMP proteins in the locus (MMP-1, MMP-7, MMP-8, MMP-10, MMP-13). Summary associations for variants in the score with CHD were obtained through PhenoScanner from a recent large-scale GWAS meta-analysis which consisted mostly (77%) of individuals of European ancestry<sup>64</sup>.

**MR analysis.** Two-sample univariable MR was performed for each protein separately using summary statistics in the inverse-variance weighted method adapted to account for correlated variants<sup>65,66</sup>. For each of  $G$  genetic variants ( $g = 1, \dots, G$ ) having per-allele estimate of the association with the protein  $\beta_{Xg}$  and standard error  $\sigma_{Xg}$ , and per-allele estimate of the association with the outcome (here, AD or CHD)  $\beta_{Yg}$  and standard error  $\sigma_{Yg}$ , the IV estimate ( $\hat{\theta}_{XY}$ ) is obtained from generalized weighted linear regression of the genetic associations with the outcome ( $\beta_Y$ ) on the genetic associations with the protein ( $\beta_X$ ) weighting for the precisions of the genetic associations with the outcome and accounting for correlations between the variants according to the regression model:

$$\beta_Y = \theta_{XY} \beta_X + \varepsilon, \quad \varepsilon \sim N(0, \Omega)$$

where  $\beta_Y$  and  $\beta_X$  are vectors of the univariable (marginal) genetic associations, and the weighting matrix  $\Omega$  has terms  $\Omega_{g_1 g_2} = \sigma_{Yg_1} \sigma_{Yg_2} \rho_{g_1 g_2}$ , and  $\rho_{g_1 g_2}$  is the correlation between the  $g_1$ th and  $g_2$ th variants.

The IV estimate from this method is:

$$\hat{\theta}_{XY} = (\beta_X^T \Omega^{-1} \beta_X)^{-1} \beta_X^T \Omega^{-1} \beta_Y$$

and the standard error is:

$$se(\hat{\theta}_{XY}) = \sqrt{(\beta_X^T \Omega^{-1} \beta_X)^{-1}}$$

where  $T$  is a matrix transpose. This is the estimate and standard error from the regression model fixing the residual standard error to 1 (equivalent to a fixed-effects model in a meta-analysis).

Genetic variants in univariable MR need to satisfy three key assumptions to be valid instruments: (1) the variant is associated with the risk factor of interest (that is, the protein level), (2) the variant is not associated with any confounder of the risk factor-outcome association, and (3) the variant is conditionally independent of the outcome given the risk factor and confounders.

To account for potential effects of functional pleiotropy<sup>67</sup>, we performed multivariable MR using the weighted regression-based method proposed by

Burgess et al.<sup>68</sup>. For each of  $K$  risk factors in the model ( $k = 1, \dots, K$ ), the weighted regression-based method is performed by multivariable generalized weighted linear regression of the association estimates  $\beta_Y$  on each of the association estimates with each risk factor  $\beta_{Xk}$  in a single regression model:

$$\beta_Y = \theta_{XY1}\beta_{X1} + \theta_{XY2}\beta_{X2} + \dots + \theta_{XYK}\beta_{XK} + \varepsilon, \varepsilon \sim N(0, \Omega)$$

where  $\beta_{X1}$  is the vectors of the univariable genetic associations with risk factor 1, and so on. This regression model is implemented by first pre-multiplying the association vectors by the Cholesky decomposition of the weighting matrix, and then applying standard linear regression to the transformed vectors. Estimates and standard errors are obtained fixing the residual standard error to be 1 as above.

The multivariable MR analysis allows the estimation of the causal effect of a protein on disease outcome accounting for the fact that genetic variants may be associated with multiple proteins in the region. Causal estimates from multivariable MR represent direct causal effects, representing the effect of intervening on one risk factor in the model while keeping others constant.

**MMP-12 genetic score sensitivity analyses.** We performed two sensitivity analyses to determine the robustness of the MR findings. First, we measured plasma MMP-12 levels using a different method (proximity extension assay; Olink Bioscience, Uppsala, Sweden<sup>13</sup>) in 4,998 individuals, and used this to derive genotype-MMP12 effect estimates for the 17 variants in our genetic score. Second, we obtained effect estimates from a pQTL study based on SOMAscan assay measurements in an independent sample of ~1,000 individuals<sup>3</sup>. In both cases the genetic score reflecting higher plasma MMP-12 was associated with lower risk of CHD.

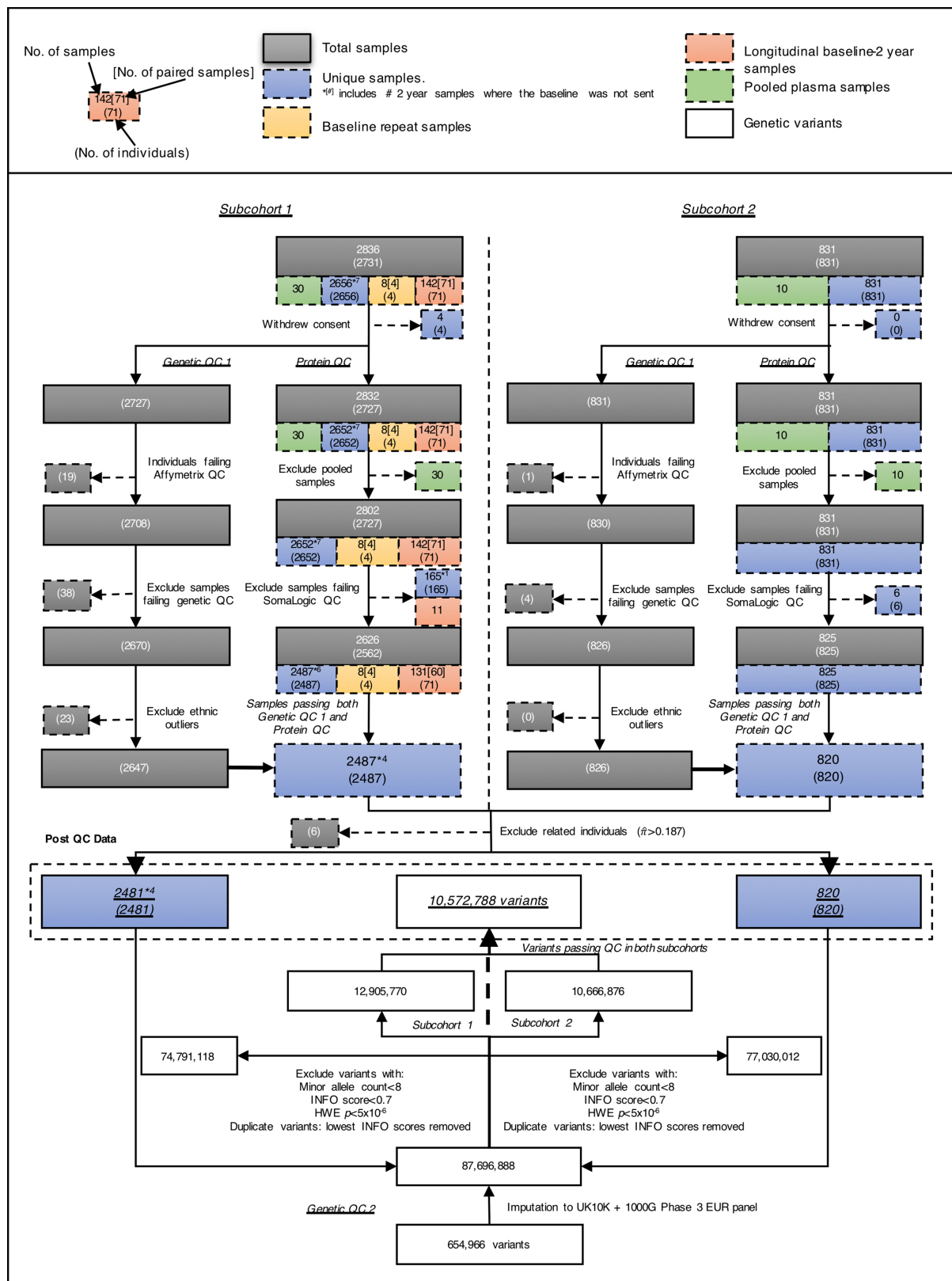
**Overlap of pQTLs with drug targets.** We used the Informa Pharmaprojects database from Cyteline to obtain information on drugs that target proteins assayed on the SOMAscan platform. This is a manually curated database that maintains profiles for >60,000 drugs. For our analysis, we focused on the following information for each drug: protein target, indications, and development status. We included drugs across the development pipeline, including those in pre-clinical studies or with no development reported, drugs in clinical trials (all phases), and launched/registered drugs. For each protein assayed, we identified all drugs in the Informa Pharmaprojects with a matching protein target based on UniProt ID. When multiple drugs targeted the same protein, we selected the drug with the latest stage of development.

For drug targets with significant pQTLs, we identified the subset where the sentinel variant or proxy variants in LD ( $r^2 > 0.8$ ) are also associated with disease risk through PhenoScanner. We used an internal Merck auto-encoding method to map GWAS traits and drug indications to a common set of terms from the Medical Dictionary for Regulatory Activities (MedDRA). MedDRA terms are organized into a hierarchy with five levels. We mapped each GWAS trait and indication onto the 'lowest level terms' (that is, the most specific terms available). All matching terms were recorded for each trait or indication. We matched GWAS traits to drug indications on the basis of the highest level of the hierarchy, called 'system organ class' (SOC). We designated a protein as 'matching' if at least one GWAS trait term matched with at least one indication term for at least one drug.

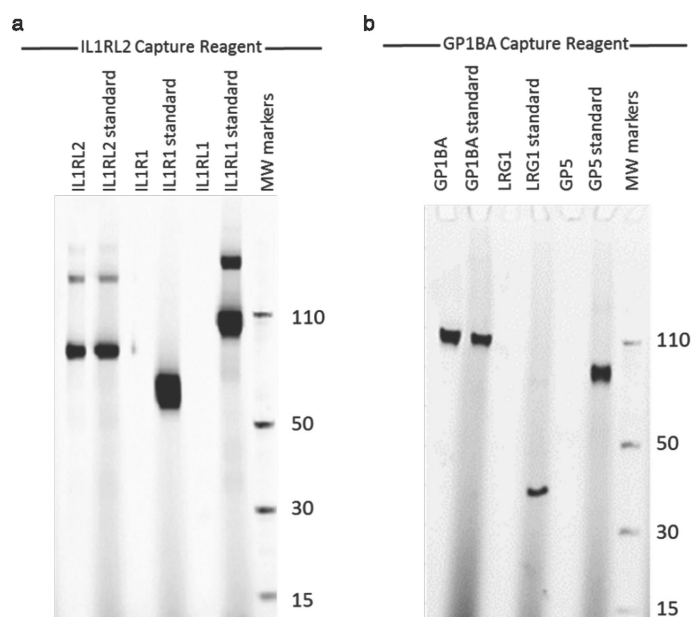
**Reporting summary.** Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

**Data availability.** Participant-level genotype and protein data, and full summary association results from the genetic analysis, are available through the European Genotype Archive (accession number EGAS00001002555). Summary association results are also publicly available at <http://www.phpc.cam.ac.uk/ceu/proteins/>, through PhenoScanner (<http://www.phenoscaner.medschl.cam.ac.uk>) and from the NHGRI-EBI GWAS Catalog (<https://www.ebi.ac.uk/gwas/downloads/summary-statistics>).

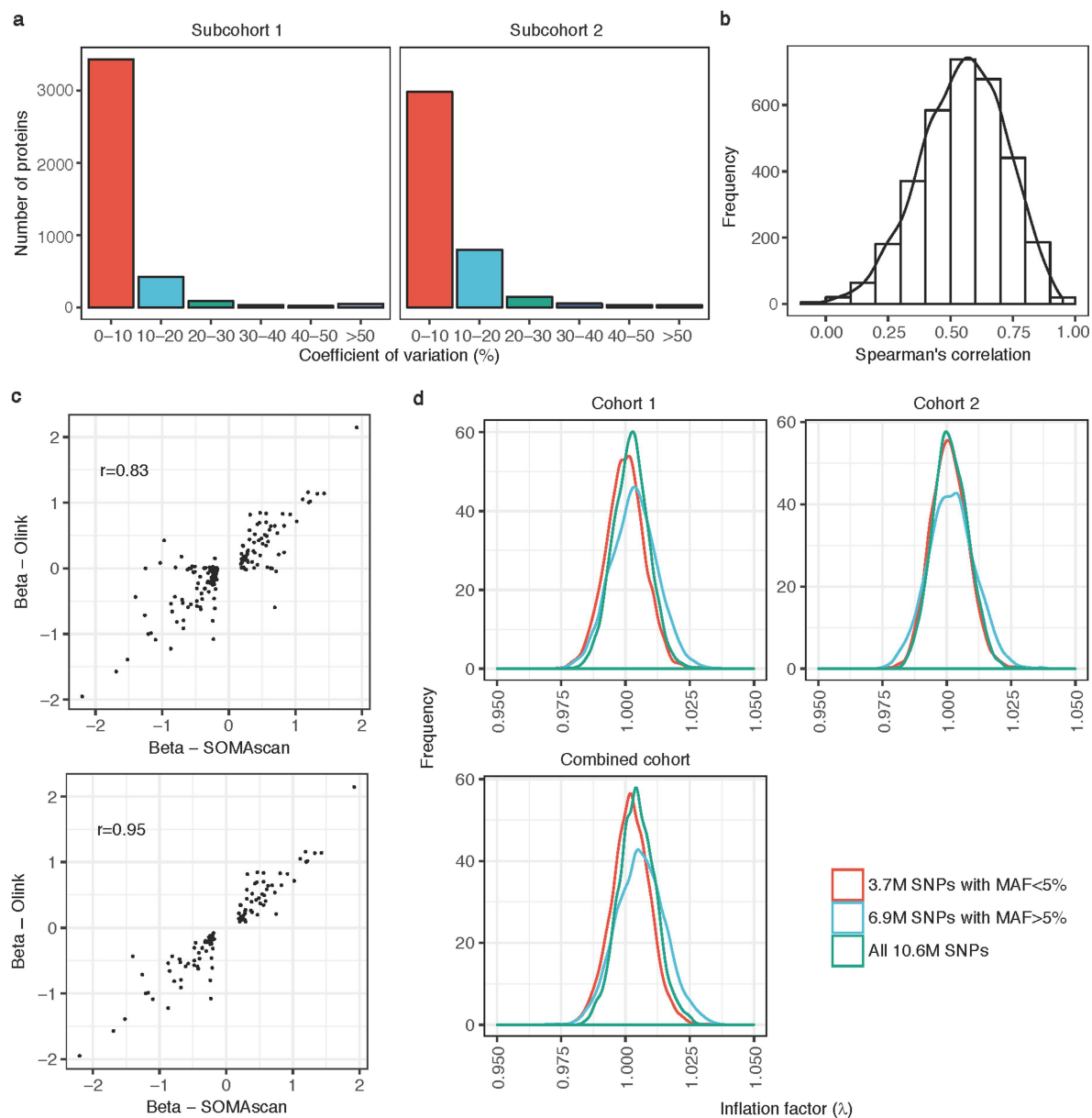
38. Moore, C. et al. The INTERVAL trial to determine whether intervals between blood donations can be safely and acceptably decreased to optimise blood supply: study protocol for a randomised controlled trial. *Trials* **15**, 363 (2014).
39. Gold, L. et al. Aptamer-based multiplexed proteomic technology for biomarker discovery. *PLoS ONE* **5**, e15004 (2010).
40. Sattlecker, M. et al. Alzheimer's disease biomarker discovery using SOMAscan multiplexed protein technology. *Alzheimers Dement.* **10**, 724–734 (2014).
41. McLaren, W. et al. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics* **26**, 2069–2070 (2010).
42. UniProt Consortium. UniProt: a hub for protein information. *Nucleic Acids Res.* **43**, D204–D212 (2015).
43. Ashburner, M. et al.; The Gene Ontology Consortium. Gene ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).
44. Menni, C. et al. Circulating proteomic signatures of chronological age. *J. Gerontol. A Biol. Sci. Med. Sci.* **70**, 809–816 (2015).
45. Ngo, D. et al. Aptamer-based proteomic profiling reveals novel candidate biomarkers and pathways in cardiovascular disease. *Circulation* **134**, 270–285 (2016).
46. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
47. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.* **39**, 906–913 (2007).
48. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
49. Yang, J. et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, 369–375 (2012).
50. Welter, D. et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* **42**, D1001–D1006 (2014).
51. Malone, J. et al. Modeling sample variables with an Experimental Factor Ontology. *Bioinformatics* **26**, 1112–1118 (2010).
52. Javierre, B. M. et al. Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. *Cell* **167**, 1369–1384. e19 (2016).
53. Lawrence, M. et al. Software for computing and annotating genomic ranges. *PLoS Comput. Biol.* **9**, e1003118 (2013).
54. Amberger, J. S., Bocchini, C. A., Schiettecatte, F., Scott, A. F. & Hamosh, A. OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. *Nucleic Acids Res.* **43**, D789–D798 (2015).
55. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* **44**, D457–D462 (2016).
56. Szklarczyk, D. et al. STRINGv10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* **43**, D447–D452 (2015).
57. Smith, R. N. et al. InterMine: a flexible data warehouse system for the integration and analysis of heterogeneous biological data. *Bioinformatics* **28**, 3163–3165 (2012).
58. Franceschini, A. et al. STRINGv9.1: protein–protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.* **41**, D808–D815 (2013).
59. Lek, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
60. Iotchkova, V. et al. GARFIELD—GWAS Analysis of Regulatory or Functional Information Enrichment with LD correction. Preprint at <https://www.biorxiv.org/content/early/2016/11/07/085738> (2016).
61. Staley, J. R. et al. PhenoScanner: a database of human genotype–phenotype associations. *Bioinformatics* **32**, 3207–3209 (2016).
62. Giambartolomei, C. et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* **10**, e1004383 (2014).
63. Hingorani, A. & Humphries, S. Nature's randomised trials. *Lancet* **366**, 1906–1908 (2005).
64. Nikpay, M. et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat. Genet.* **47**, 1121–1130 (2015).
65. Burgess, S., Butterworth, A. & Thompson, S. G. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet. Epidemiol.* **37**, 658–665 (2013).
66. Burgess, S., Dudbridge, F. & Thompson, S. G. Combining information on multiple instrumental variables in Mendelian randomization: comparison of allele score and summarized data methods. *Stat. Med.* **35**, 1880–1906 (2016).
67. Burgess, S. & Thompson, S. G. Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. *Am. J. Epidemiol.* **181**, 251–260 (2015).
68. Burgess, S., Dudbridge, F. & Thompson, S. G. Re: "Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects". *Am. J. Epidemiol.* **181**, 290–291 (2015).



Extended Data Fig. 1 | Flowchart of sample processing and quality control stages for proteomic and genetic measurements before genetic analyses.

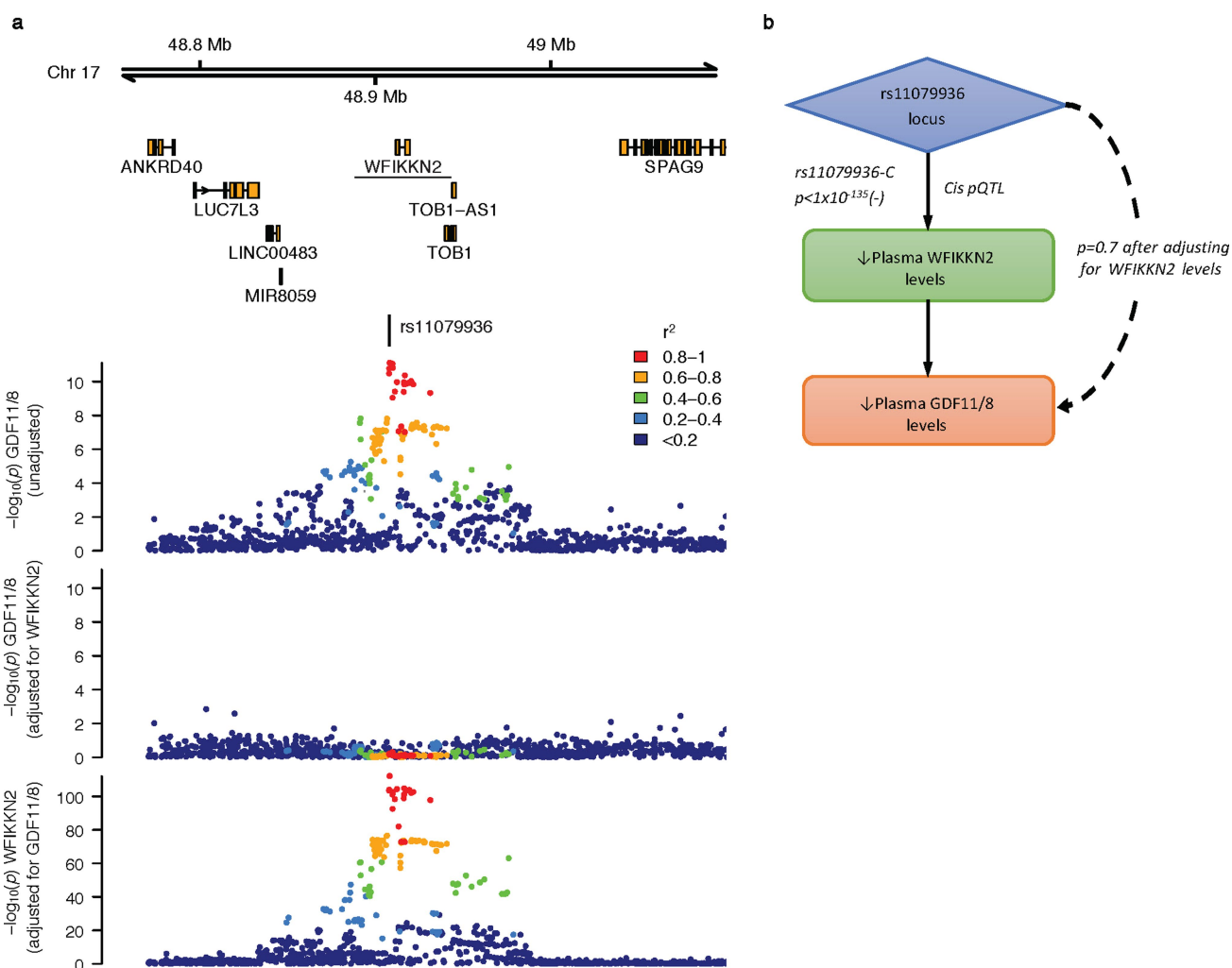


**Extended Data Fig. 2 | Examples of protein targets for which the SOMAmer is highly specific.** SDS-PAGE with Alexa-647-labelled proteins captured by the IL1RL2 SOMAmer (a) or GP1BA SOMAmer (b). For each protein target, the protein captured by the SOMAmer is compared to the standard. The cognate targets are the only ones with protein visible in the capture lanes, whereas the proteins homologous to the target proteins show no evidence of binding. These experiments were performed once. MW markers, molecular weight markers.



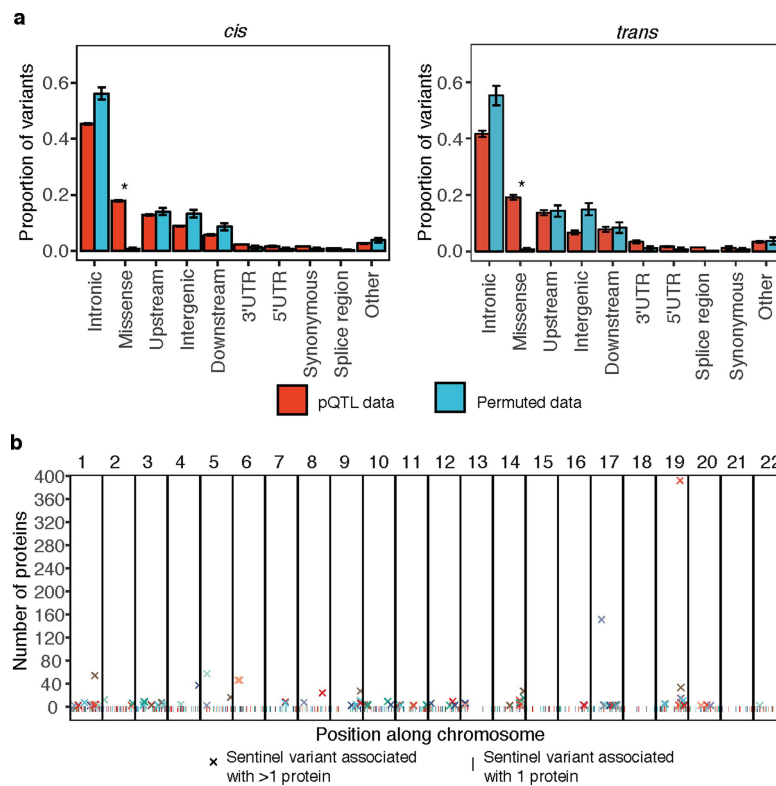
**Extended Data Fig. 3 | Evidence for the reliability of protein measurements made using the SOMAscan assay.** **a**, Distribution of coefficients of variation of all proteins on the SOMAscan assay in each subcohort. **b**, Spearman's correlations for all proteins passing QC derived from contemporaneous assay of baseline and two-year samples from 60

participants. **c**, Scatterplot of pQTL effect size estimates from SOMAscan versus Olink showing all 163 pQTLs tested (top) and the 106 that replicated (bottom).  $r$  is Pearson's correlation coefficient. **d**, Distribution of inflation factors across proteins that underwent genome-wide association testing, stratified by subcohort and allele frequency (MAF  $\geq$  5%, MAF < 5%).



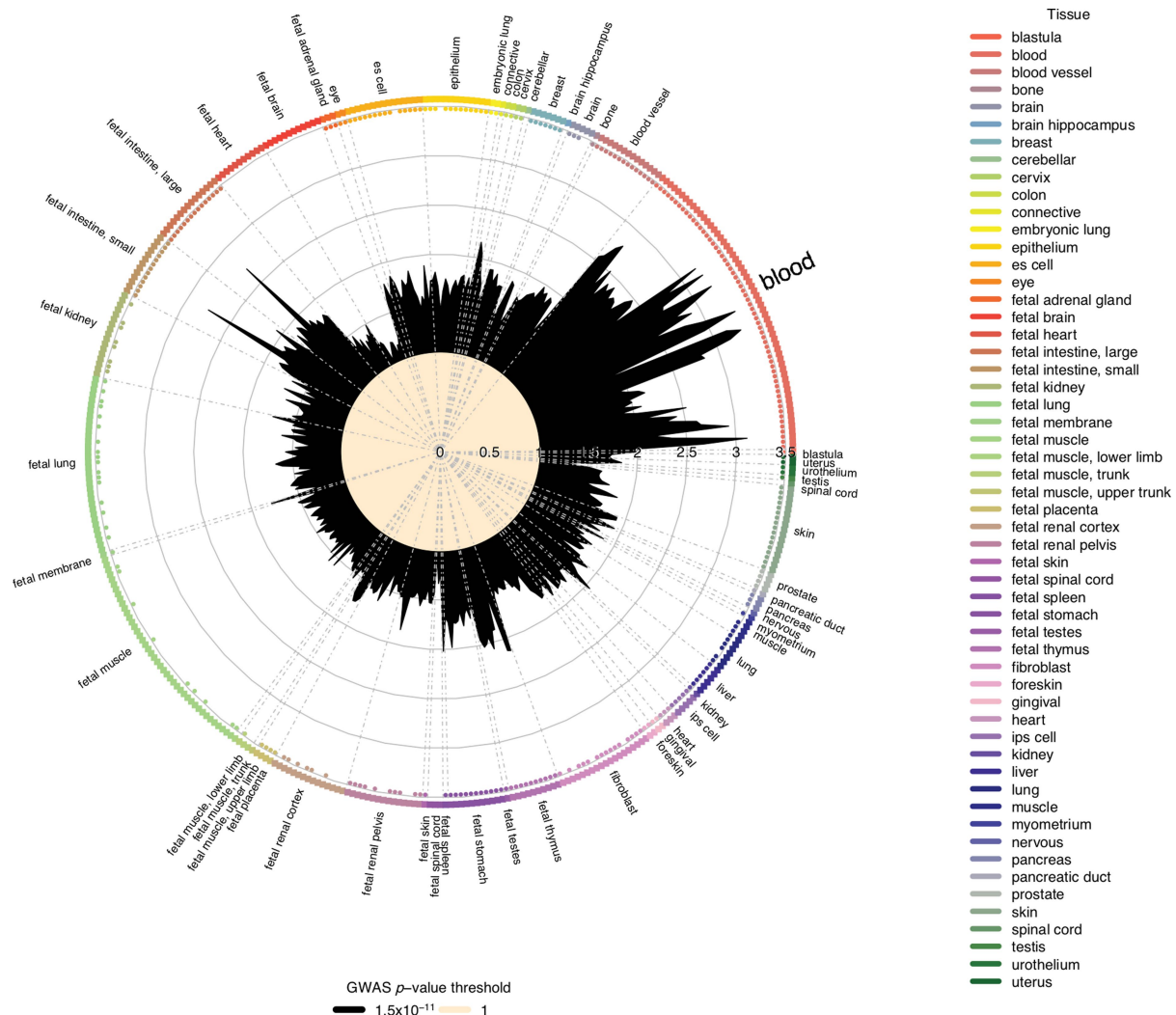
**Extended Data Fig. 4 | The *WFIKK2* region is a *trans* pQTL for GDF11/8 plasma levels. a**, Regional association plots of the *trans* pQTL (sentinel variant rs11079936) for GDF11/8 before and after adjusting for levels of WFIKK2 (upper panels), and the WFIKK2 *cis* pQTL

after adjusting for GDF11/8 levels (bottom panel). A similar pattern of association for WFIKK2 was seen before GDF11/8 adjustment (not shown). **b**, Attenuation of the GDF11/8 *trans* pQTL upon adjustment for plasma levels of the *cis* protein WFIKK2.



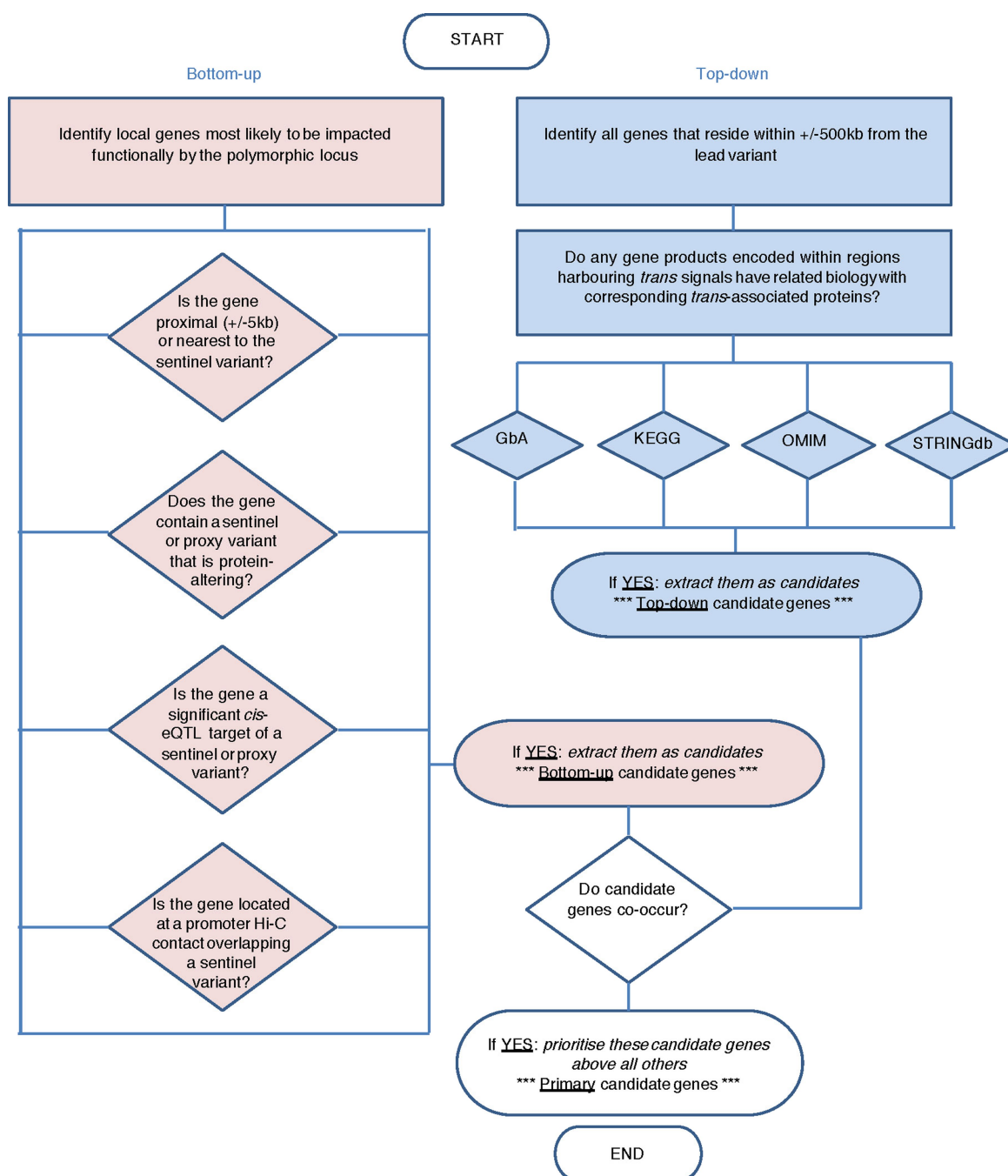
**Extended Data Fig. 5 | Genetic architecture of the pQTLs.** pQTL mapping in  $n = 3,301$  individuals. **a**, Distribution of the predicted consequences of the sentinel pQTL variants compared to matched permuted null sets of variants, stratified by *cis* and *trans*. Asterisks indicate empirical enrichment using a permutation test (10,000 permuted sets

of non-associated variants) at a Bonferroni-corrected significance value ( $P < 0.005$ ). Bar height represents the mean proportion of variants within each class and error bars reflect one standard deviation from the mean. **b**, Number of proteins associated ( $P < 1.5 \times 10^{-11}$ ) with each sentinel variant across the genome.



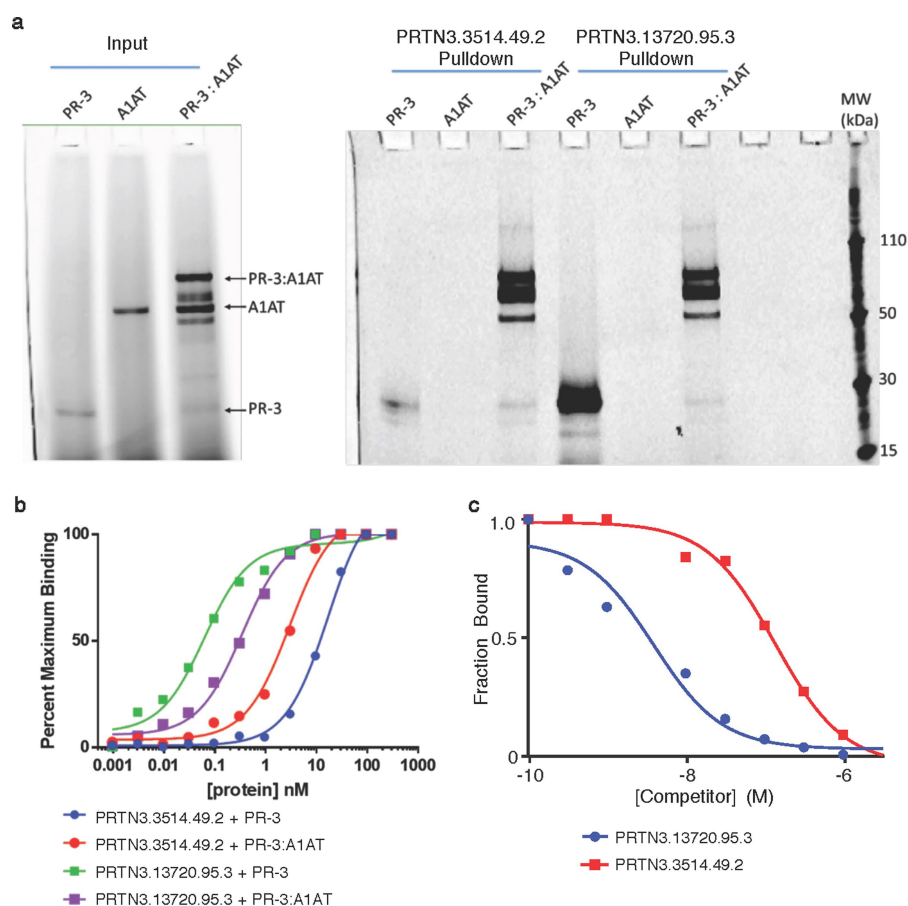
**Extended Data Fig. 6 | Enrichment of pQTLs at DNase I hypersensitive sites by tissue or cell type.** Circle shows enrichment for DNase I hypersensitive sites ('hotspots') for each of 55 tissues (183 cell types) available from the ENCODE and Roadmap Epigenomics projects, with tissues or cell types clustered and coloured by anatomical grouping. Some tissues have multiple values due to availability of multiple cell types or

multiple tests per cell type. Radial lines show fold-enrichment, while dots around the inside edge of the circle denote statistically significant enrichment at a Bonferroni-corrected significant threshold  $P < 5 \times 10^{-5}$ . Enrichment testing performed using GARFIELD (which tests enrichment against permuted sets of variants matched for MAF, distance to TSS and LD). pQTL data from  $n = 3,301$  individuals.



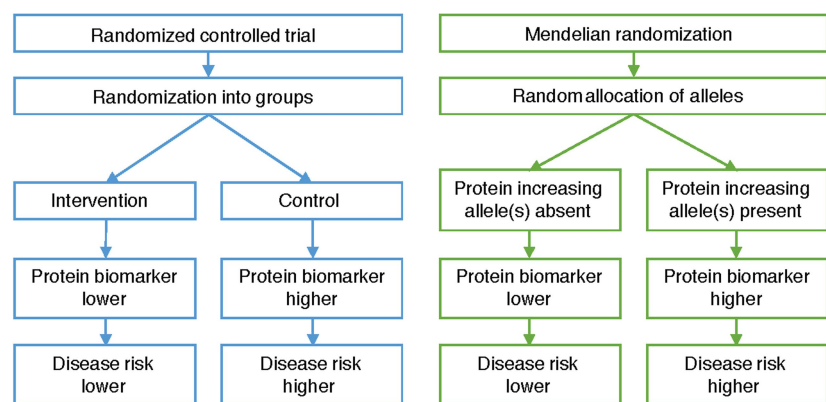
**Extended Data Fig. 7 | Scheme outlining the combined 'bottom-up' and 'top-down' process used for candidate gene annotation of *trans* pQTL regions.** See Methods. GbA, guilt-by-association; KEGG, Kyoto

Encyclopedia of Genes and Genomes; OMIM, Online Mendelian Inheritance in Man; STRINGdb, STRING database.

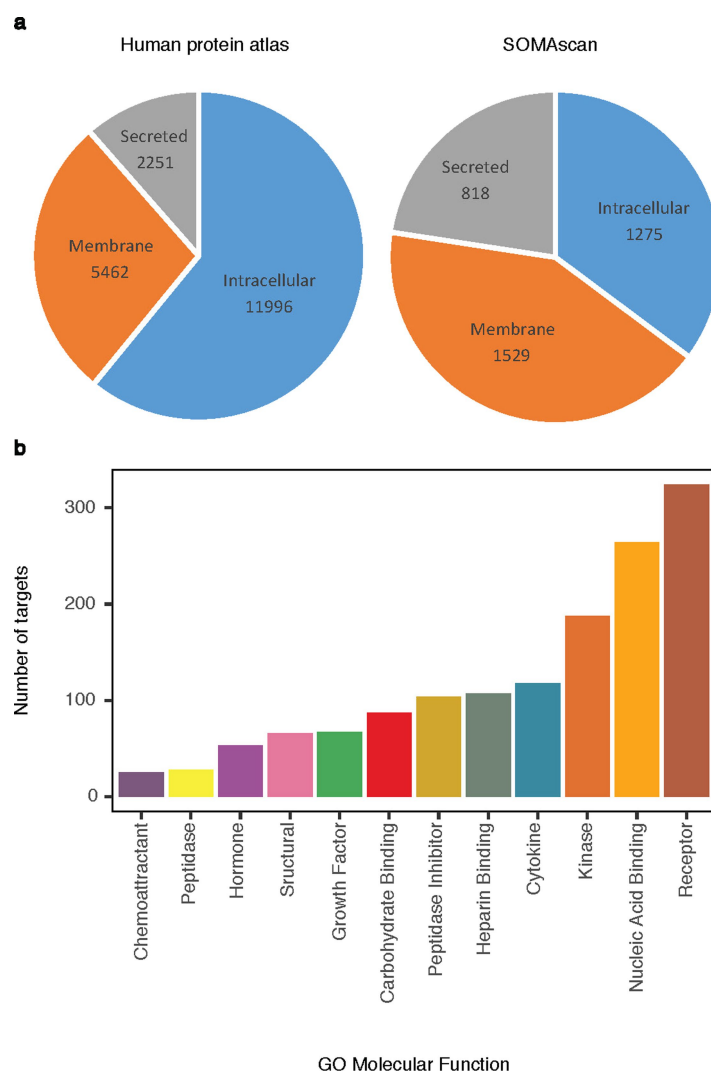


**Extended Data Fig. 8 | Follow-up of PR3 SOMAmers.** These experiments were repeated three times independently with similar results. **a**, SOMAmer pull-downs with purified PR3, A1AT, and PR3–A1AT complex. SOMAmer PRTN3.3514.49.2 enriched the PR3–A1AT complex to a much greater degree than free PR3. Conversely, SOMAmer PRTN3.13720.95.3 enriched free PR3 to a greater degree than the PR3–A1AT complex. **b**, Solution affinity of PRTN3.3514.49.2 and PRTN3.13720.95.3 for PR3,

A1AT, and the PR3–A1AT complex. SOMAmer PRTN3.3514.49.2 has a higher affinity for the PR3–A1AT complex than for free PR3. SOMAmer PRTN3.13720.95.3, on the other hand, has a higher affinity for free PR3 than SOMAmer PRTN3.3514.49.2. **c**, Competitive binding of SOMAmers PRTN3.13720.95.3 and PRTN3.3514.49.2 to PR3. A limiting amount of radiolabelled PRTN3.13720.95.3 was incubated with 1 nM proteinase-3 and a titration of either cold PRTN3.13720.95.3 or cold PRTN3.3514.49.2.



Extended Data Fig. 9 | Comparison between a randomized controlled trial and Mendelian randomization to assess the causal effect of changes in protein biomarker levels on disease risk.



**Extended Data Fig. 10 | Characterization of protein targets measured using the SOMAscan assay.** **a**, Compartment distribution with annotations of all proteins in the Human Protein Atlas for comparison. **b**, GO molecular functions.

# Cortical direction selectivity emerges at convergence of thalamic synapses

Anthony D. Lien<sup>1,2,3\*</sup> & Massimo Scanziani<sup>1,2,3,4\*</sup>

**Detecting the direction of motion of an object is essential for our representation of the visual environment. The visual cortex is one of the main stages in the mammalian nervous system in which the direction of motion may be computed de novo. Experiments and theories indicate that cortical neurons respond selectively to motion direction by combining inputs that provide information about distinct spatial locations with distinct time delays. Despite the importance of this spatiotemporal offset for direction selectivity, its origin and cellular mechanisms are not fully understood. We show that approximately  $80 \pm 10$  thalamic neurons, which respond with distinct time courses to stimuli in distinct locations, excite mouse visual cortical neurons during visual stimulation. The integration of thalamic inputs with the appropriate spatiotemporal offset provides cortical neurons with a primordial bias for direction selectivity. These data show how cortical neurons selectively combine the spatiotemporal response diversity of thalamic neurons to extract fundamental features of the visual world.**

Detecting the direction of motion of visual stimuli is an essential part of sensory processing. In the mammalian primary visual cortex (V1), many neurons show direction selectivity, preferentially responding to stimuli that move in a specific direction<sup>1,2</sup>. In primates and carnivores, it is thought that direction selectivity is computed de novo in V1<sup>1,3</sup>. V1 neurons probably extract directional motion by combining inputs that respond with different temporal delays to stimuli in different spatial locations<sup>3–11</sup>. Most models of direction selectivity, from the most general<sup>12,13</sup> to those based on V1 receptive fields<sup>14–17</sup>, rely on this spatiotemporal offset (although see ref. 18). However, the cellular and synaptic mechanisms that generate this spatiotemporal offset remain speculative.

Models propose both intracortical and thalamocortical synapses as potential sources of spatiotemporal offset. In intracortical models (Fig. 1a; model 1), anisotropic connectivity between neurons results in spatial offset between excitation and inhibition or between excitatory inputs with distinct time courses<sup>3,15,16,19–23</sup>. In thalamocortical models (Fig. 1a; model 2), direction selectivity emerges from the synaptic convergence of spatially and temporally offset thalamic inputs onto cortical neurons<sup>8,24–27</sup>. Alternative models suggest that V1 inherits direction selectivity from the retina<sup>28–31</sup> (Fig. 1a; model 3). Here we isolate individual thalamic inputs onto mouse layer 4 (L4) cortical neurons to identify a mechanism for the de novo generation of direction selectivity in V1.

## Thalamic excitation reports motion direction

We performed whole-cell patch-clamp recordings from V1 L4 neurons<sup>32</sup> (300–550  $\mu\text{m}$  from the pia; Fig. 1b). Visual stimuli consisted of gratings of six orientations drifting in either one of two opposite directions. When presented with gratings in their preferred orientation, some L4 neurons (Fig. 1c) fired more in response to movement in one direction than in the other (Fig. 1d), consistent with previous studies<sup>33</sup>. The membrane potential ( $V_m$ ) fluctuated at the temporal frequency (2 Hz) of the drifting grating (F1 modulation; Fig. 1c, d). The direction selectivity index (DSI; Methods) of the amplitude of the F1 modulation of  $V_m$  (F1  $V_m$ ) correlated with the DSI of the spiking response of the

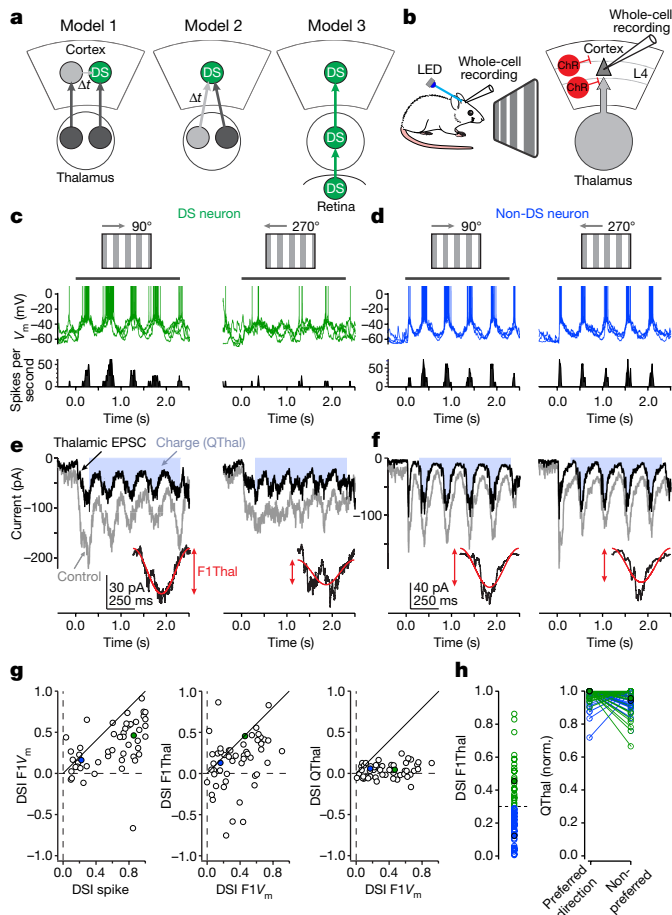
neuron ( $r = 0.52$ ;  $P = 0.000103$ ; Fig. 1g, left). In most cells, the preferred direction was the same for both parameters (46 out of 51;  $P < 0.0001$ ; binomial test; Fig. 1g left).

To determine whether direction selectivity is already apparent in the thalamic input, we isolated the thalamic excitation. Cortical excitatory neurons were silenced by photoactivation of cortical inhibitory interneurons expressing channelrhodopsin-2 (ChR2)<sup>34</sup> while recordings from L4 neurons were obtained in the voltage-clamp configuration at the reversal potential of inhibition<sup>32</sup> (Fig. 1b, e, f, Extended Data Fig. 1).

We analysed thalamic excitatory currents using two parameters: the amplitude of F1 modulation (F1Thal) and the charge (QThal), that is, the time integral of the current (Fig. 1e, f). F1Thal showed direction selectivity in 38% of the neurons (that is, the DSI of F1Thal was greater than 0.3 in 25 out of 66 neurons; Fig. 1h, left). The preferred direction of F1Thal matched that of F1  $V_m$  in the majority of neurons (37 out of 52;  $P = 0.0032$ ; binomial test; Fig. 1g, middle). This was also the case for neurons in which the DSI of F1  $V_m$  was larger than 0.3 (22 out of 28;  $P = 0.0037$ , binomial test). The DSI values of F1Thal and F1  $V_m$  were correlated ( $r = 0.33$ ;  $P = 0.0168$ ;  $n = 52$ ; Fig. 1g, middle).

By contrast, QThal showed no direction selectivity (that is, DSI  $< 0.3$  in 64 out of 66 neurons) and was similar for the preferred and non-preferred directions of F1Thal (for DSI F1Thal  $> 0.3$ , the normalized QThal values were  $0.98 \pm 0.03$  in the preferred and  $0.93 \pm 0.09$  in the non-preferred direction;  $P = 0.13$ ;  $n = 25$ ; for DSI F1Thal  $< 0.3$ , the normalized QThal values were  $0.97 \pm 0.06$  preferred,  $0.95 \pm 0.06$  non-preferred;  $P = 0.22$ ;  $n = 41$ ; Wilcoxon rank-sum tests; Fig. 1h, right). DSI QThal and DSI F1  $V_m$  were not correlated ( $r = 0.2$ ;  $P = 0.16$ ;  $n = 52$  cells; Fig. 1g, right) and QThal did not predict the preferred direction of F1  $V_m$  (32 out of 52 preferred the same direction;  $P = 0.126$ ; binomial test; for DSI F1  $V_m > 0.3$ , 18 out of 28 preferred the same direction;  $P = 0.1849$ ; binomial test; Fig. 1g, right). Nevertheless, when the preferred direction of QThal matched that of F1  $V_m$  (positive points in Fig. 1g, right) the average absolute value of DSI QThal was slightly, yet significantly, larger than when they did not match (negative points in Fig. 1g, right; Extended Data Fig. 2). To quantify the contribution of QThal to DSI F1Thal, assuming linearity, we arithmetically equalized

<sup>1</sup>Neurosciences Graduate Program, University of California San Diego, La Jolla, CA, USA. <sup>2</sup>Center for Neural Circuits and Behavior, Neurobiology Section and Department of Neuroscience, University of California San Diego, La Jolla, CA, USA. <sup>3</sup>Department of Physiology, University of California San Francisco, San Francisco, CA, USA. <sup>4</sup>Howard Hughes Medical Institute, University of California San Francisco, San Francisco, CA, USA. \*e-mail: anthony.d.lien@gmail.com; massimo@ucsf.edu



**Fig. 1 | Amplitude modulation of thalamic excitation is direction selective.** **a**, Models of the direction selectivity of V1. DS, direction-selective neuron;  $\Delta t$ , temporal delay. **b**, Recording configuration for the isolation of thalamic excitation. **c**, Top, membrane potential ( $V_m$ ) of a direction-selective neuron in response to gratings drifting in the preferred (left) and non-preferred (right) direction (four sweeps). Bottom, the PSTH. The horizontal bar indicates the duration of the visual stimulus. **d**, As in **c**, but for a non-direction-selective neuron. **e**, Voltage-clamp recording ( $V_{\text{holding}} = -70$  mV) of the neuron in **c** under control (grey) and cortical silencing (black) conditions to isolate thalamic excitation. The inset shows the cycle average and the sinusoidal fit (red). The red arrow indicates the amplitude of modulation ( $F1Thal$ ); blue shading indicates the thalamic excitatory charge ( $QThal$ ). **f**, As in **e**, but for the neuron in **d**. **g**, Summary plots ( $n = 52$ ) of the relationship between DSI spike, DSI  $F1V_m$ , DSI  $F1Thal$  and DSI  $QThal$ . Negative y-axis values indicate the opposite direction preference to the index of the x axis. The diagonal indicates unity. **h**, Left, distribution of DSI  $F1Thal$  ( $n = 66$ ). DSI  $F1Thal$  values greater than 0.3 (dotted line) are considered to be direction selective. Right,  $QThal$  (normalized to its preferred direction) in the preferred and non-preferred direction of  $F1Thal$ . In **g**, **h**, the filled points indicate the neurons in **c**, **e** (green) and in **d**, **f** (blue).

$QThal$  across directions. This caused a small, yet significant reduction in DSI  $F1Thal$  (for DSI  $F1Thal > 0.3$ , the DSI was  $0.49 \pm 0.15$  before and  $0.46 \pm 0.17$  after equalization;  $P = 0.022$ ; paired  $t$ -test;  $n = 25$ ; Extended Data Fig. 2; Methods). These data demonstrate that direction selectivity in L4 neurons is already prominent in the amplitude modulation of thalamic excitation<sup>27</sup>, but much less so in the charge. As such, essentially equal amounts of thalamic excitation are differently distributed in time depending on the direction of the stimulus.

### Time course of thalamic excitation

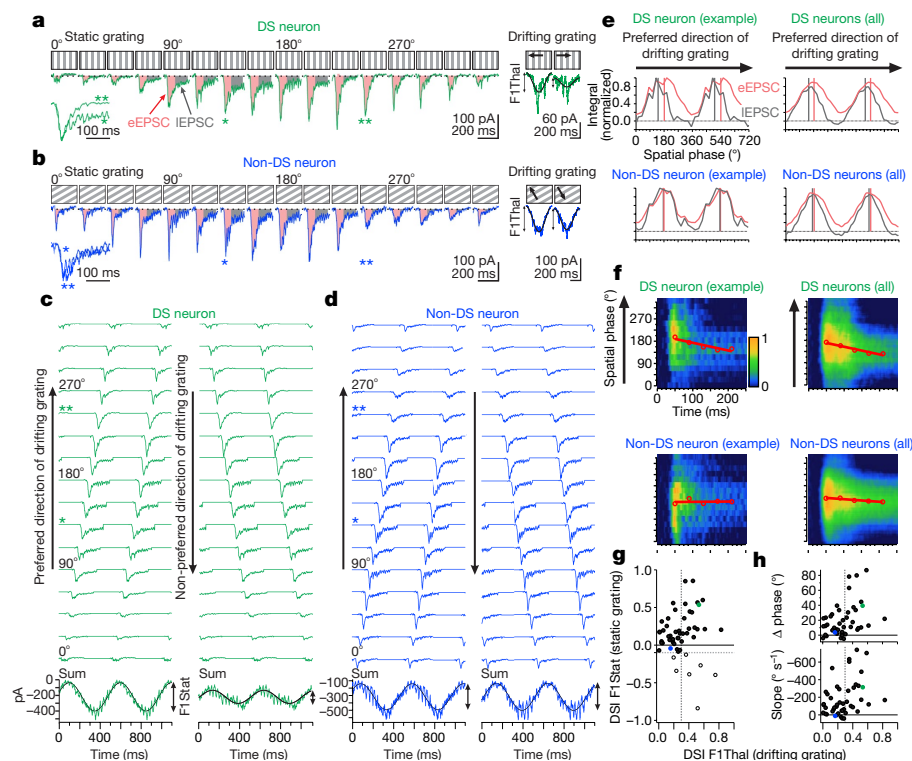
We investigated the mechanisms for the differential temporal distribution of thalamic excitation for stimuli moving in opposite directions. A possible mechanism is that the direction-selective

amplitude modulation results from a heterogeneity in the time course of thalamic excitation elicited by stimuli at different receptive field locations<sup>8,24–27</sup>.

To determine whether the time course of thalamic excitation depends on the position of the stimulus, we presented—while silencing the cortex—static gratings of the preferred orientation at 16 different spatial phases (Fig. 2a, b).

First we validated that static gratings recapitulate the dynamics of thalamic excitation in response to drifting gratings. We computed the algebraic sum of thalamic excitatory postsynaptic currents (EPSCs) evoked by each of the 16 phases of the static gratings (Fig. 2c, d; Extended Data Table 1) appropriately staggered in time, to simulate gratings drifting at 2 Hz in the preferred and non-preferred directions (Methods; Extended Data Fig. 3 for the simulation of additional temporal frequencies). In most L4 neurons, the  $F1$  modulation of the summed thalamic EPSCs evoked by static gratings ( $F1Stat$ ) showed preference for the same direction as  $F1Thal$  (41 out of 53;  $P < 0.0001$ ; binomial test; Fig. 2g). Neurons for which DSI  $F1Thal$  is greater than 0.3 showed similar results (18 out of 23;  $P = 0.01$ ; binomial test). Among the 12 mismatches, 7 were not direction selective (DSI  $F1Thal < 0.3$ ; Fig. 2g). Although we have no explanation for the mismatched directional preference of the remaining five cells, their  $QThal$  to drifting gratings was on average not direction selective (absolute value of DSI  $QThal$   $0.09 \pm 0.07$ ;  $n = 5$ ), consistent with the data above. Without those five cells, DSI  $F1Stat$  and DSI  $F1Thal$  were correlated ( $r = 0.4$ ;  $P = 0.0041$ ;  $n = 48$ ), but they were not correlated when these five cells were included ( $r = 0.043$ ;  $P = 0.76$ ;  $n = 53$ ). These results validate static gratings for determining the dynamics of thalamic excitation that underlie direction selectivity. Below we restrict our analysis to those 46 neurons for which DSI  $F1Stat$  is greater than  $-0.1$  (Fig. 2g).

Static gratings evoked thalamic EPSCs, the amplitude of which depended on the spatial phase, consistent with spatially offset ON and OFF sub-regions of thalamic excitation<sup>32</sup> (Fig. 2a, asterisks). We analysed the decay of thalamic EPSCs by measuring the integral of the early (Fig. 2a, b, pink) and late (Fig. 2a, b, grey) portion of the EPSC (Fig. 2e). The integral of the early and late EPSC fluctuated sinusoidally with a period spanning the 16 phases of the static gratings. Crucially, the phase difference between the integral of the early and late EPSC varied between cells (Fig. 2e), ranging from  $0^\circ$  (that is, the same preferred spatial phase for early and late EPSCs) to  $87^\circ$  (that is, the preferred spatial phases for the early and late EPSCs are shifted relative to each other; Fig. 2h, top). This phase difference predicted the direction preference of the cell to drifting gratings. The predicted preferred direction was that in which the preferred spatial phase of the late EPSC preceded that of the early EPSC. Intuitively, two consecutive EPSCs with distinct decays summate to a larger peak current if the slow one precedes the fast. In 89% of neurons (41 out of 46), the predicted preferred direction matched that observed with drifting gratings ( $P < 0.0001$ ; binomial test). The phase difference between early and late EPSCs correlated with DSI  $F1Thal$  ( $r = 0.405$ ;  $P = 0.0053$ ;  $n = 46$ ) and was significantly larger in direction-selective neurons (DSI  $F1Thal < 0.3$ :  $10.8 \pm 11.5^\circ$ ;  $n = 28$ ; DSI  $F1Thal > 0.3$ :  $33.8 \pm 26.9^\circ$ ;  $n = 18$ ;  $P = 0.000603$ ; Wilcoxon rank-sum test; Fig. 2h, top). To visualize the phase difference we plotted the spatiotemporal receptive field, a heat map in which the time course of the EPSC is shown for each spatial phase of the static grating (Fig. 2f). We identified the preferred phase of each time bin and fitted a linear function. Steeper slopes indicate larger phase differences between early and late EPSCs. Negative slopes predict a preferred direction towards increasing spatial phase. Again, in 89% of neurons (41 out of 46), the predicted preferred direction matched that observed with drifting gratings ( $P < 0.0001$ ; binomial test). The slope correlated with DSI  $F1Thal$  ( $r = -0.38$ ;  $P = 0.009$ ;  $n = 46$ ) and was significantly steeper for direction-selective neurons (DSI  $F1Thal < 0.3$ :  $-104 \pm 129^\circ/s$ ;  $n = 28$ ; DSI  $F1Thal > 0.3$ :  $-311 \pm 232^\circ/s$ ;  $n = 18$ ;  $P = 0.00204$ ; Wilcoxon rank-sum test; Fig. 2h, bottom). Therefore,



**Fig. 2 | The time course of thalamic excitation to static stimuli explains direction selectivity to moving stimuli.** **a**, Thalamic EPSCs recorded in a direction-selective neuron (DSI F1Thal = 0.53). Left, static grating response to 16 different spatial phases (0–337.5°, 250-ms duration). Early and late time integral shaded in pink (eEPSC; 30–100 ms after stimulus onset) and grey (IEPSC; 110–230 ms), respectively. Left inset, superimposed, peak-scaled thalamic EPSCs elicited by two different phases (asterisks). Note the difference in decay. Right, drifting grating cycle average and sinusoidal fit (black) in preferred (left) and non-preferred (right) direction. **b**, As in **a**, but for a non-direction-selective neuron (DSI F1Thal = 0.16). **c**, Responses to static gratings (from **a**) staggered in time (top) and their sum (bottom) simulating the response to 2.2 cycles of gratings drifting in the preferred (left) and non-preferred (right) direction. Asterisks indicate the EPSCs shown in **a**. The black trace is the sinusoidal fit showing amplitude modulation (F1Stat). **d**, As in **c**, but for the non-direction-selective neuron in **b**. **e**, Phase-dependent modulation of eEPSC and IEPSC. Top and bottom left, an example direction-selective and non-direction-selective neuron from **a** and **b**, respectively. The cycle is duplicated for clarity. Right, average for

direction-selective ( $n = 18$ ) and non-direction-selective ( $n = 28$ ) neurons. The preferred spatial phase of eEPSC (vertical bars) was set at 180°. Note the phase-shift of IEPSC relative to eEPSC in direction-selective but not in non-direction-selective neurons. **f**, Spatiotemporal receptive field of thalamic EPSCs to static gratings. The preferred phase at each time bin and the linear fit are shown in red. Top and bottom left, example neurons in **a** and **b**, respectively; right, the average for the same neurons as in **e**. Population heat maps were shifted so that the preferred spatial phase of the earliest time bin was 180°. The arrow indicates the preferred direction to drifting gratings. Note the different slopes. **g**, DSI to drifting gratings (DSI F1Thal) plotted against DSI of summed responses to static gratings (DSI F1Stat) ( $n = 53$ ). Only seven neurons (indicated by empty circles below the horizontal dotted line, and not included in further analysis) had a strong mismatch between drifting and static direction preference (DSI F1Stat < −0.1). The blue circles indicate the example neurons in **c** and **d**. **h**, Phase difference between eEPSC and IEPSC (top; see **e**) and slope of linear fits (bottom; see **f**) versus DSI F1Thal ( $n = 46$ ). The vertical dotted line indicates DSI = 0.3.

different stimulus positions trigger thalamic excitation with different early and late components, which provides an initial bias for direction preference in L4 neurons.

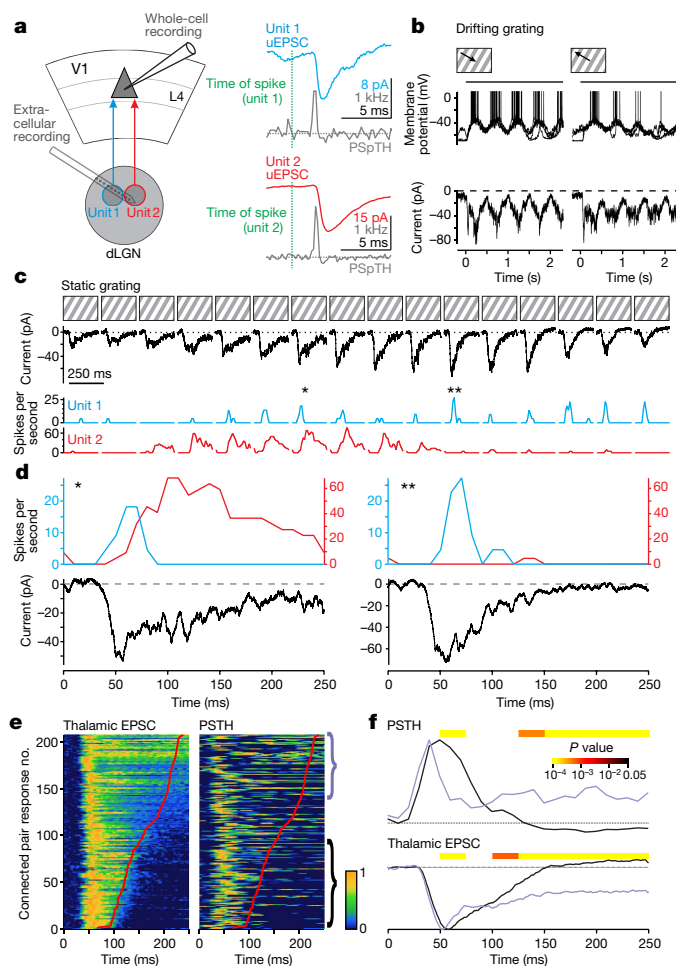
### Time course of single thalamic inputs

We determined whether the phase dependence of the time course of firing of presynaptic thalamic neurons by recording from synaptically connected thalamocortical pairs<sup>35,36</sup>. We simultaneously recorded thalamic single units in the dorsal lateral geniculate nucleus (dLGN) and excitation from L4 neurons (Fig. 3a). Across the population, the response of thalamic units to static gratings varied from transient to sustained<sup>37</sup> (Extended Data Fig. 4). For individual units, the response type (transient or sustained) was not modulated by the phase of the grating; however, the firing rate was (Extended Data Fig. 4).

Twenty-three thalamocortical monosynaptic connections were identified on the basis of the latency, time course and probability (captured by the perispikes time histogram) of EPSCs in the cortical neuron following thalamic unit spikes (Fig. 3a; Methods and Extended Data Figs. 5, 6; see Extended Data Table 2 for the properties of unitary EPSCs (uEPSCs)).

The response to static gratings of presynaptic thalamic units matched the duration of thalamic EPSCs in postsynaptic L4 neurons. Figure 3 shows two thalamic units—unit 1 showing transient responses and unit 2 showing sustained responses—converging on the same direction-selective L4 neuron (Fig. 3a, b). These units responded maximally to distinct spatial phases (Fig. 3c, bottom). Phases that drove transient spiking in unit 1 and little spiking in unit 2 elicited fast-decaying EPSCs in the L4 neuron (Fig. 3d, left), whereas phases that drove sustained spiking in unit 2 elicited slow-decaying EPSCs (Fig. 3d, right).

To compare the time course of thalamic EPSCs in L4 neurons with the spiking of their presynaptic thalamic units, we created a heat map of EPSCs ranked by duration (Fig. 3e, left; Methods). Every row shows the time course of the thalamic EPSC recorded in one L4 neuron in response to one spatial phase of the static grating. Corresponding rows in the adjacent heat map show the peristimulus time histogram (PSTH) of spiking in a presynaptic thalamic unit (Fig. 3e, right). There was a marked and significant correlation between the time course of thalamic EPSCs in L4 neurons and the spiking of their presynaptic thalamic units (average Pearson's correlation 0.40;  $n = 208$  paired EPSC and PSTH responses; significantly different than shuffled;  $P < 0.0001$ ; Methods). Accordingly, the average PSTH of presynaptic thalamic

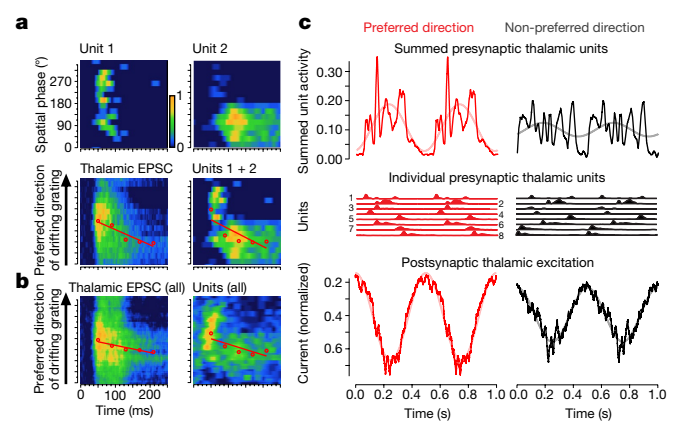


**Fig. 3 | Time course of thalamic spiking explains the time course of thalamic excitation.** **a**, Left, recording configuration. Right, two uEPSCs (blue and red) from two presynaptic thalamic units (1 and 2) recorded in the same L4 neuron while silencing the cortex. The green line indicates the time of the thalamic spike; the grey traces are the perispoke time histogram (PSPTH). The horizontal dotted line shows an event frequency of 0.25 kHz. **b**, Direction-selective response of the L4 neuron in **a** to gratings drifting in preferred (left) and non-preferred (right) directions. Top, current-clamp recording (truncated spikes) and bottom, voltage-clamp recording of thalamic excitation during cortical silencing. The horizontal bars indicate the duration of the stimuli. **c**, Responses of neurons in **a** to static gratings at the preferred orientation of the L4 neuron. Top, thalamic excitation; bottom, PSTHs. **d**, Responses to two spatial phases (indicated by the asterisks in **c**) on expanded axes. **e**, Heat maps of static grating responses for 20 connected pairs sorted by duration (red line) of thalamic EPSC ( $n = 208$  responses). Left, each row shows the thalamic EPSC for one spatial phase. Right, the PSTH of the presynaptic thalamic unit appears on the corresponding row. Slower thalamic excitation is accompanied by more sustained PSTHs (upper portion of heat maps). **f**, Top, average PSTHs from the upper (purple) and lower (black) rows (brackets) of the right heat map in **e**. Bottom, average thalamic EPSC for the same rows.  $P$  values indicated are for the Wilcoxon rank-sum test comparing black and purple traces in 20-ms bins.

units was significantly more sustained for spatial phases that triggered slow-decaying thalamic EPSCs compared with those that triggered fast-decaying EPSCs (Fig. 3f). Therefore, the time course of thalamic excitation in an L4 neuron depends on spatial phase and follows the spiking time course of its presynaptic thalamic units.

### Combining individual thalamic inputs

Given the above results, L4 neurons may extract motion direction by combining thalamic inputs with distinct spatial and temporal response properties. As such, the spatiotemporal receptive field of thalamic

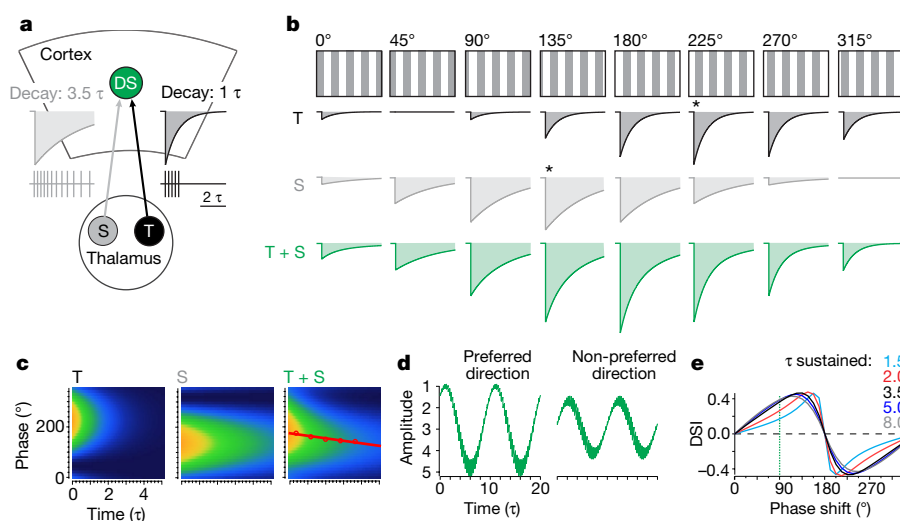


**Fig. 4 | Convergence of spatiotemporally distinct thalamic units generates direction selectivity.** **a**, Spatiotemporal receptive fields of thalamic units 1 and 2, their sum, and thalamic excitation (thalamic EPSC) of their postsynaptic L4 neuron (same recording as Fig. 3a–d). The preferred spatial phase at each time bin and the linear fit are shown in red. **b**, Receptive field of the thalamic EPSC in the compound neuron (left; average of six direction-selective L4 neurons) and its eight presynaptic thalamic units (right; two L4 neurons received input from two recorded thalamic units; for example, Fig. 3a). The preferred spatial phase at each time bin and the linear fit are shown in red. **c**, Drifting grating response of the compound neuron and its inputs. Top, summed activity of presynaptic thalamic units in the preferred (red) and non-preferred (black) directions of the compound neuron (cycle duplicated for clarity; pink and grey lines indicate sinusoidal fits). Middle, the PSTHs of each unit (units 1 and 2 correspond to those in **a**). Bottom, thalamic excitation of the compound neuron.

excitation of an L4 neuron (for example, Fig. 2f) should reflect the combined spatiotemporal receptive field of its presynaptic thalamic neurons. Indeed, the combined spatiotemporal receptive field of units 1 and 2 (from example above; Fig. 3a–d) approximates that of their postsynaptic L4 neuron (Fig. 4a), with sustained unit 2 preferring lower spatial phases and transient unit 1 preferring higher spatial phases. Because recordings of several thalamic units converging on the recorded L4 neuron were rare, we confirmed this observation by generating a ‘compound’ direction-selective cortical neuron and its thalamic inputs. We pooled responses from all eight connected pairs with cortical DSI F1Thal > 0.3 (six L4 and eight thalamic neurons; Fig. 4b). The spatiotemporal receptive field of the compound cortical neuron (Fig. 4b; Methods) had a negative slope, consistent with its six constituent neurons preferring gratings drifting in the direction of increasing spatial phase. The compound spatiotemporal receptive field of thalamic inputs also had a negative slope: whereas the more sustained units were active at the lower spatial phases, transient thalamic units dominated the higher spatial phases. Therefore, the combined spatiotemporal receptive field of presynaptic thalamic units predicted the direction preference of the compound postsynaptic neuron (Fig. 4b).

We analysed the response of the direction-selective compound neuron and its thalamic inputs to drifting gratings (Fig. 4c; Methods). In the preferred direction of the L4 neuron, the thalamic neurons fired in phase to produce strongly F1-modulated population spiking (Fig. 4c). In the opposite direction, firing was more evenly distributed over time, which resulted in weakly F1-modulated population spiking. Furthermore, consistent with the de novo emergence of direction selectivity in the cortex, the DSI of the firing of thalamic neurons and the DSI of their postsynaptic L4 neuron were uncorrelated (Extended Data Fig. 7a). Indeed, arithmetically removing the directional bias in thalamic firing had no effect on the results (Extended Data Fig. 7b). Therefore, the convergence of transient and sustained thalamic neurons preferring distinct spatial phases produces a spatiotemporal offset that confers direction preference to the postsynaptic L4 neuron.

A model (Fig. 5) in which two converging thalamic neurons—a transient and a sustained—prefer distinct spatial phases (Fig. 5a) captures



**Fig. 5 | A simple model of direction selectivity.** **a**, Sustained (S) and transient (T) thalamic neurons with spatially offset receptive fields converge on a direction-selective cortical neuron, generating slow- (light grey;  $3.5\tau$ ) and fast- (dark grey;  $1\tau$ ) decaying EPSCs. **b**, EPSCs mediated by T and S neurons and their sum (T + S; green) in response to static gratings of various spatial phases. Owing to receptive field offset (here by  $90^\circ$  in phase space; asterisks show the largest response), the relative contribution of fast- and slow-decaying EPSCs to the decay of the compound EPSC (T + S) varies with spatial phase. **c**, Spatiotemporal receptive field of EPSCs generated by T and S neurons and their sum (T +

S). The preferred spatial phase at each time bin and the linear fit are shown in red. The spatiotemporal receptive field of the sum is tilted. **d**, Thalamic excitation in the cortical neuron for gratings drifting in the preferred (left) and non-preferred (right) direction. Cycle period is  $10\tau$ . **e**, DSI of F1 modulation of thalamic excitation versus spatial phase difference (phase shift) of T and S neurons for the above example (black) and for different decays of the slow EPSC (coloured traces). The vertical green line marks the  $90^\circ$  phase-shift in the above example. For phase shifts of  $0^\circ$  or  $180^\circ$ , the relative contribution of S and T inputs is identical for all phases and thalamic excitation lacks direction selectivity (DSI = 0).

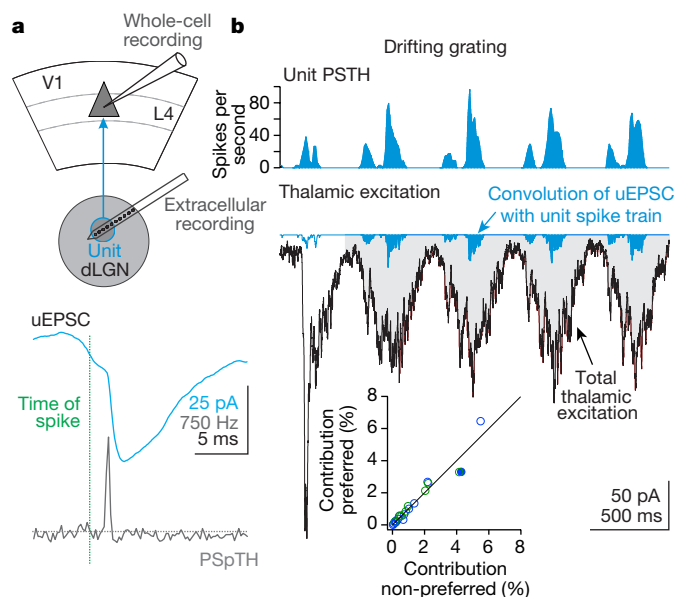
the above observations (Fig. 5). The transient and sustained neurons generate a fast- and a slow-decaying EPSC, respectively, which combine into a compound EPSC (Fig. 5b). Because the thalamic neurons prefer distinct spatial phases, the decay of the compound EPSC changes with phase (Fig. 5b, c). Time-staggered summation of compound EPSCs results in large or small F1 modulation depending on the direction of simulated motion (Fig. 5d). The direction preference and the DSI of the cortical neuron depend on the phase shift between the two thalamic neurons (Fig. 5e).

Finally, using the 23 pairs, we estimated the number of thalamic neurons that excite an L4 neuron during drifting gratings (Fig. 6). We calculated a unitary contribution for each thalamic neuron by convolving its uEPSC and spiking response. The mean unitary contribution was  $1.25 \pm 1.51\%$ , indicating that approximately  $80 \pm 10$  thalamic neurons excite an L4 neuron (Fig. 6b; Extended Data Fig. 8), independently of stimulus direction ( $1.25 \pm 1.54\%$  preferred;  $1.25 \pm 1.51\%$  non-preferred;  $P = 0.52$ ; Wilcoxon signed-rank test).

## Discussion

These results show that direction selectivity in L4 neurons originates from the combination of thalamic inputs with distinct spatiotemporal response properties, consistent with model 2 (Fig. 1a). Intracortical mechanisms may enhance direction selectivity through anisotropic inhibition or excitation<sup>3,15,16,19–22</sup>, recurrent excitation<sup>15,32,38</sup> and intrinsic excitable properties of the neuronal membrane, including dendrites<sup>11,39</sup>. By contrast, because QThal was poorly direction selective (Fig. 1g, h) and the direction preference of cortical neurons did not correlate with that of their thalamic inputs (Extended Data Fig. 7), direction selectivity in the retina has a minor role in L4 direction selectivity. However, we cannot exclude the possibility that other mechanisms contribute to the emergence of direction selectivity and may account for those neurons in which direction selectivity was not predicted by static gratings responses (Fig. 2g), and that anaesthesia may have impaired those mechanisms.

Classical models based on spatiotemporal offsets in excitation<sup>12</sup> or inhibition relative to excitation<sup>13</sup> underlie most of the proposed mechanisms of cortical direction selectivity<sup>3,15,16,19–21</sup>. The spatiotemporal



**Fig. 6 | Contribution of individual thalamic neurons to thalamic excitation in the visual cortex.** **a**, Top, recording configuration. Bottom, uEPSC recorded in an L4 neuron. The time of the thalamic unit spike is indicated by the green line, and the grey trace shows the PSpTH of postsynaptic events. **b**, Top, PSTH of the thalamic unit during drifting gratings while silencing the cortex. Middle, thalamic excitatory current (black) recorded in the L4 neuron and estimated unitary thalamic excitation (blue indicates convolution of uEPSC with unit spike train). The shading indicates the time integral of thalamic excitation. The contribution of the thalamic unit was approximately 4% of the total thalamic excitation. Bottom, unitary contributions in preferred and non-preferred directions for all pairs (green, 8 with DSI  $F1_{Thal} > 0.3$ ; blue, 15 with DSI  $F1_{Thal} < 0.3$ ). The filled blue circle indicates the example neuron above, and the diagonal indicates unity.

offset of excitatory inputs, originally formulated by Reichardt<sup>12</sup>, more closely captures the mechanism we describe.

dLGN neurons with different response time courses—categorized as transient and sustained<sup>40</sup> or lagged and non-lagged<sup>24,41</sup>—exist in cats<sup>24,40,41</sup>, monkeys<sup>42</sup> and mice<sup>37</sup>, and have been proposed to provide the spatiotemporal offset for cortical direction selectivity<sup>24,25</sup>. Although our model (Fig. 5) considers only dLGN neurons with identical onsets, late onset neurons were recorded (Extended Data Fig. 4) and might enhance cortical direction selectivity.

We may have underestimated average uEPSC amplitudes, and hence overestimated the number of thalamic neurons that excite an L4 neuron (Extended Data Fig. 8), because large uEPSCs were probably undersampled and submillisecond synchrony between thalamic units might cause erroneous identification of monosynaptic connections. Nevertheless, our estimate of the amplitude of unitary excitatory postsynaptic potential (Extended Data Fig. 6; Extended Data Table 2) is similar to that of previous paired recordings in the rodent somatosensory cortex<sup>35,43</sup> and the cat visual cortex<sup>36</sup>, and our estimate of thalamic convergence onto L4 neurons (approximately 80 neurons) is similar to estimates in the somatosensory cortex<sup>35</sup>.

We conclude that, in the mammalian nervous system, direction selectivity is generated de novo in at least two stages of visual processing: the retina and the cortex. Studies in rodents indicate that some dLGN neurons inherit direction selectivity from the retina<sup>28,37,44</sup> and project to V1<sup>29,30</sup>. However, the abolition of retinal direction selectivity does not eliminate V1 direction selectivity<sup>31</sup>, consistent with its de novo emergence in the cortex. Whether direction selectivity computed in the cortex is combined, at some stage, with direction selectivity from the retina remains to be established.

## Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0148-5>.

Received: 10 July 2017; Accepted: 16 April 2018;

Published online: 23 May 2018

- Hubel, D. H. & Wiesel, T. N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol. (Lond.)* **160**, 106–154 (1962).
- Hubel, D. H. & Wiesel, T. N. Receptive fields of single neurones in the cat's striate cortex. *J. Physiol. (Lond.)* **148**, 574–591 (1959).
- Livingstone, M. S. Mechanisms of direction selectivity in macaque V1. *Neuron* **20**, 509–526 (1998).
- Movshon, J. A., Thompson, I. D. & Tolhurst, D. J. Spatial summation in the receptive fields of simple cells in the cat's striate cortex. *J. Physiol. (Lond.)* **283**, 53–77 (1978).
- McLean, J. & Palmer, L. A. Contribution of linear spatiotemporal receptive field structure to velocity selectivity of simple cells in area 17 of cat. *Vision Res.* **29**, 675–679 (1989).
- Albrecht, D. G. & Geisler, W. S. Motion selectivity and the contrast-response function of simple cells in the visual cortex. *Vis. Neurosci.* **7**, 531–546 (1991).
- Reid, R. C., Soodak, R. E. & Shapley, R. M. Directional selectivity and spatiotemporal structure of receptive fields of simple cells in cat striate cortex. *J. Neurophysiol.* **66**, 505–529 (1991).
- Saul, A. B. & Humphrey, A. L. Evidence of input from lagged cells in the lateral geniculate nucleus to simple cells in cortical area 17 of the cat. *J. Neurophysiol.* **68**, 1190–1208 (1992).
- DeAngelis, G. C., Ohzawa, I. & Freeman, R. D. Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. II. Linearity of temporal and spatial summation. *J. Neurophysiol.* **69**, 1118–1135 (1993).
- Jagadeesh, B., Wheat, H. S., Kontsevich, L. L., Tyler, C. W. & Ferster, D. Direction selectivity of synaptic potentials in simple cells of the cat visual cortex. *J. Neurophysiol.* **78**, 2772–2789 (1997).
- Priebe, N. J. & Ferster, D. Direction selectivity of excitation and inhibition in simple cells of the cat primary visual cortex. *Neuron* **45**, 133–145 (2005).
- Reichardt, W. in *Sensory Communication* (ed. Rosenblith, W. A.) 303–317 (MIT Press, Cambridge MA, 1961).
- Barlow, H. B. & Levick, W. R. The mechanism of directionally selective units in rabbit's retina. *J. Physiol. (Lond.)* **178**, 477–504 (1965).
- Adelson, E. H. & Bergen, J. R. Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Am. A* **2**, 284–299 (1985).
- Suarez, H., Koch, C. & Douglas, R. Modeling direction selectivity of simple cells in striate visual cortex within the framework of the canonical microcircuit. *J. Neurosci.* **15**, 6700–6719 (1995).
- Maex, R. & Orban, G. A. Model circuit of spiking neurons generating directional selectivity in simple cells. *J. Neurophysiol.* **75**, 1515–1545 (1996).
- Watson, A. B. & Ahumada, A. J. Jr. Model of human visual-motion sensing. *J. Opt. Soc. Am. A* **2**, 322–341 (1985).
- Li, Y. T., Fang, Q., Zhang, L. I. & Tao, H. W. Spatial asymmetry and short-term suppression underlie direction selectivity of synaptic excitation in the mouse visual cortex. *Cereb. Cortex* <https://doi.org/10.1093/cercor/bhx111> (2017).
- Goodwin, A. W., Henry, G. H. & Bishop, P. O. Direction selectivity of simple striate cells: properties and mechanism. *J. Neurophysiol.* **38**, 1500–1523 (1975).
- Emerson, R. C. & Gerstein, G. L. Simple striate neurons in the cat. II. Mechanisms underlying directional asymmetry and directional selectivity. *J. Neurophysiol.* **40**, 136–155 (1977).
- Ganz, L. & Felder, R. Mechanism of directional selectivity in simple neurons of the cat's visual cortex analyzed with stationary flash sequences. *J. Neurophysiol.* **51**, 294–324 (1984).
- Li, Y. T., Liu, B. H., Chou, X. L., Zhang, L. I. & Tao, H. W. Strengthening of direction selectivity by broadly tuned and spatiotemporally slightly offset inhibition in mouse visual cortex. *Cereb. Cortex* **25**, 2466–2477 (2015).
- Sillito, A. M. Inhibitory processes underlying the directional specificity of simple, complex and hypercomplex cells in the cat's visual cortex. *J. Physiol. (Lond.)* **271**, 699–720 (1977).
- Saul, A. B. & Humphrey, A. L. Spatial and temporal response properties of lagged and nonlagged cells in cat lateral geniculate nucleus. *J. Neurophysiol.* **64**, 206–224 (1990).
- Alonso, J.-M., Usrey, W. M. & Reid, R. C. Rules of connectivity between geniculate cells and simple cells in cat primary visual cortex. *J. Neurosci.* **21**, 4002–4015 (2001).
- Stanley, G. B. et al. Visual orientation and directional selectivity through thalamic synchrony. *J. Neurosci.* **32**, 9073–9088 (2012).
- Ferster, D., Chung, S. & Wheat, H. Orientation selectivity of thalamic input to simple cells of cat visual cortex. *Nature* **380**, 249–252 (1996).
- Marshall, J. H., Kaye, A. P., Nauhaus, I. & Callaway, E. M. Anterior–posterior direction opponency in the superficial mouse lateral geniculate nucleus. *Neuron* **76**, 713–720 (2012).
- Cruz-Martin, A. et al. A dedicated circuit links direction-selective retinal ganglion cells to the primary visual cortex. *Nature* **507**, 358–361 (2014).
- Sun, W., Tan, Z., Mensh, B. D. & Ji, N. Thalamus provides layer 4 of primary visual cortex with orientation- and direction-tuned inputs. *Nat. Neurosci.* **19**, 308–315 (2016).
- Hillier, D. et al. Causal evidence for retina-dependent and -independent visual motion computations in mouse cortex. *Nat. Neurosci.* **20**, 960–968 (2017).
- Lien, A. D. & Scanziani, M. Tuned thalamic excitation is amplified by visual cortical circuits. *Nat. Neurosci.* **16**, 1315–1323 (2013).
- Niell, C. M. & Stryker, M. P. Highly selective receptive fields in mouse visual cortex. *J. Neurosci.* **28**, 7520–7536 (2008).
- Boyden, E. S., Zhang, F., Bamberg, E., Nagel, G. & Deisseroth, K. Millisecond-timescale, genetically targeted optical control of neural activity. *Nat. Neurosci.* **8**, 1263–1268 (2005).
- Bruno, R. M. & Sakmann, B. Cortex is driven by weak but synchronously active thalamocortical synapses. *Science* **312**, 1622–1627 (2006).
- Sedigh-Sarvestani, M. et al. Intracellular, in vivo, dynamics of thalamocortical synapses in visual cortex. *J. Neurosci.* **37**, 5250–5262 (2017).
- Piscopo, D. M., El-Danaf, R. N., Huberman, A. D. & Niell, C. M. Diverse visual features encoded in mouse lateral geniculate nucleus. *J. Neurosci.* **33**, 4642–4656 (2013).
- Li, Y. T., Ibrahim, L. A., Liu, B. H., Zhang, L. I. & Tao, H. W. Linear transformation of thalamocortical input by intracortical excitation. *Nat. Neurosci.* **16**, 1324–1330 (2013).
- Smith, S. L., Smith, I. T., Branco, T. & Häusser, M. Dendritic spikes enhance stimulus selectivity in cortical neurons in vivo. *Nature* **503**, 115–120 (2013).
- Cleland, B. G., Dubin, M. W. & Levick, W. R. Sustained and transient neurones in the cat's retina and lateral geniculate nucleus. *J. Physiol. (Lond.)* **217**, 473–496 (1971).
- Mastroratte, D. N. Two classes of single-input X-cells in cat lateral geniculate nucleus. I. Receptive-field properties and classification of cells. *J. Neurophysiol.* **57**, 357–380 (1987).
- Marrocco, R. T. Sustained and transient cells in monkey lateral geniculate nucleus: conduction velocities and response properties. *J. Neurophysiol.* **39**, 340–353 (1976).
- Hu, H. & Agmon, A. Differential excitation of distally versus proximally targeting cortical interneurons by unitary thalamocortical bursts. *J. Neurosci.* **36**, 6906–6916 (2016).
- Zhao, X., Chen, H., Liu, X. & Cang, J. Orientation-selective responses in the mouse lateral geniculate nucleus. *J. Neurosci.* **33**, 12751–12763 (2013).

**Acknowledgements** We thank J. Evora for help with genotyping and mouse husbandry, S. Hestrin for allowing us to perform some of the experiments in his laboratory, R. Beltramo for helping with extracellular recordings, J. S. Isaacson, B. L. Bloodgood and R. A. Nicoll for comments on the manuscript, E. J. Chichilnisky and members of the Scanziani and Isaacson laboratories

for discussions of this project. This research was supported by the Gatsby Charitable Foundation, the Howard Hughes Medical Institute and the Jane Coffin Childs Fellowship to A.D.L.

**Reviewer information** *Nature* thanks J. Alonso, S. Hofer and B. Roska for their contribution to the peer review of this work.

**Author contributions** A.D.L. and M.S. designed the study. A.D.L. conducted all experiments and analysis. A.D.L. and M.S. wrote the paper.

**Competing interests** The authors declare no competing interests.

#### Additional information

**Extended data** is available for this paper at <https://doi.org/10.1038/s41586-018-0148-5>.

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41586-018-0148-5>.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>.

**Correspondence and requests for materials** should be addressed to A.D.L. or M.S.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## METHODS

**Mice.** Experiments were performed in accordance with the regulations of the Institutional Animal Care and Use Committee of the University of California San Diego and of the University of California San Francisco and the Administrative Panel on Laboratory Animal Care at Stanford University. Mice were heterozygous male or female offspring of PV-Cre<sup>45</sup> (JAX #008069) or VGAT-ChR2<sup>46</sup> (JAX #014548) transgenic mice crossed with ICR white wild-type animals. Data were obtained from 41 VGAT-ChR2 mice and 7 PV-Cre mice. The data presented in this study were obtained from a total of 66 whole-cell recordings in V1, 7 extracellular recordings in V1 and 24 extracellular recordings in the dLGN. The number of cells and mice for each experiment are detailed here and in the main text:

V1 cells recorded in voltage clamp while showing drifting gratings (Fig. 1; Extended Data Fig. 1, left): 61 cells in 36 VGAT-ChR2 mice; 5 cells in 5 PV-Cre mice.

V1 cells recorded in both current clamp and voltage clamp while showing drifting gratings (Fig. 1): 52 cells in 31 VGAT-ChR2 mice.

V1 cells recorded in voltage clamp while showing drifting and static gratings (Fig. 2): 48 cells in 29 VGAT-ChR2 mice; 5 cells in 5 PV-Cre mice.

V1 cells recorded in voltage clamp while performing extracellular recordings in the dLGN and showing drifting and static gratings (Figs. 3, 4, 6; Extended Data Fig. 6, 7, 8): 40 cells and 739 dLGN isolated units in 24 VGAT-ChR2 mice. Out of these 739 units, 23 were considered presynaptic to the recorded V1 neuron (see 'Identification of monosynaptic connections'). Furthermore, out of these 739 units, data from 177 are reported in Extended Data Fig. 4. These 177 units were selected on the basis of their responsiveness to static gratings (see 'dLGN unit analysis').

V1 extracellular recordings (Extended Data Fig. 1): 13 units in 2 PV-Cre mice; 138 units in 5 VGAT-ChR2 mice.

dLGN extracellular recording (Extended Data Fig. 4): 739 units in 24 VGAT-ChR2 mice.

**Visual stimuli.** Visual stimuli were presented on an LCD monitor (75 cd m<sup>-2</sup> mean luminance, gamma corrected) to the eye contralateral to the hemisphere in which recordings were performed. Drifting grating stimuli were full-field, full-contrast drifting bar gratings (spatial frequency of 0.04 cycles per degree, temporal frequency of 2 Hz). For determining the preferred orientation, 12 different drifting grating stimuli were presented consisting of 6 evenly spaced orientations (30° increment) drifting in one of two opposite directions along the axis perpendicular to the grating bars. Drifting gratings were presented for 2.3 s and were preceded and followed by a grey screen of mean luminance. Presentation of a single grating was considered as one visual stimulus trial (see 'Cortical silencing methods'). In 52 out of 66 recordings, the preferred orientation and direction of the cell were determined in current clamp using the full set of 12 orientations and directions, followed by recording in the voltage-clamp configuration using drifting grating stimuli restricted to the 2 opposite directions at the preferred orientation. Although in these cells the preferred orientation of thalamic excitation recorded during cortical silencing was not determined, it probably matches the preferred orientation obtained in current clamp<sup>27,38</sup>. In the remaining 14 cells the recording was started in the voltage-clamp configuration and the preferred orientation was determined in voltage clamp, while isolating thalamic excitation during cortical silencing, by presenting the full set of 12 orientations and directions. The orientation and/or direction of the grating stimuli were presented in random order.

Static grating stimuli were full-field, full-contrast bar gratings (0.04 cycles per degree of spatial frequency) of the preferred orientation of the cortical neuron. Sixteen evenly-spaced spatial phases were presented (22.5° increments, equal to 1/16 of a cycle). A series of five static gratings of randomly chosen spatial phase was presented sequentially. Each static grating was presented for 250 ms and followed by a grey screen of mean luminance for 250 ms. Each five-grating sequence was considered as one visual stimulation trial (see 'Cortical silencing methods') and was preceded and followed by a grey screen of mean luminance.

**Surgery.** AAV1-Flex-ChR2-tdTomato (University of Pennsylvania Vector Core) was injected in the visual cortex of neonatal (postnatal day 0–1) PV-Cre mice to achieve ChR2 expression in PV<sup>+</sup> inhibitory interneurons (PV-ChR2 mice) as previously described<sup>32</sup>. VGAT-ChR2 mice express ChR2 in cortical inhibitory interneuron populations, therefore no viral injections were required.

For recording, adult mice (5–12 weeks) were anaesthetized with a combination of urethane (1.5 g kg<sup>-1</sup>, i.p.), chlorprothixene (2–4 mg kg<sup>-1</sup>, i.p.) and light isoflurane (0.5% in O<sub>2</sub>) for the duration of the experiment. A drop of silicone oil was applied to the eyes. The scalp was removed, the skull cleaned and a metal head-fixation bar was affixed to the skull using dental acrylic. A 1–2 mm diameter craniotomy was performed over V1 in one hemisphere (1 mm anterior of the lambdoid suture, 2.5 mm lateral to the midline). In simultaneous V1 and dLGN recording experiments, a second narrow elongated craniotomy was performed (spanning 2–3 mm posterior of bregma, 3.4 mm lateral to the midline) for insertion of the silicon probe array into the dLGN. In both craniotomies the dura was removed. Recording began shortly after completion of surgery.

**Cortical silencing.** The visual cortex was silenced by illuminating the V1 craniotomy with a 1-mm fibre optic coupled to a blue LED (470 nm, 20 mW total output, Doric) positioned several millimetres above the craniotomy, or through the objective (20×) of a fluorescence microscope with a blue LED (470 nm, 2.3 mW total output, Thorlabs) coupled to the excitation port<sup>32</sup>. The illumination may even extend beyond the borders of V1 but this was not measured. The LED was turned on 650 ms before the onset of a visual stimulus trial and lasted throughout the duration of the visual stimulus. Trials with cortical silencing were interleaved with trials without illumination in which cortical activity was intact. To validate the effectiveness of cortical silencing, spiking responses of V1 neurons to drifting gratings were recorded using loose-patch or silicon probes (Buzsaki32 or A1x32-Edge-5mm-20-177, NeuroNexus). Cortical silencing in both PV-ChR2 and VGAT-ChR2 mice suppressed nearly all spiking of non-narrow spiking neurons in V1 (approximately 99% suppression, Extended Data Fig. 1b) and reduced visually evoked excitation by around 65% (Extended Data Fig. 1h), consistent with the fact that, during visual stimulation, the majority of excitation received by L4 neurons is of cortical origin<sup>27,32,38</sup>. Most known excitatory inputs to L4 in V1 originate from either within V1 or from the dLGN, and therefore we consider visually evoked excitation isolated during cortical silencing to originate from the dLGN. However, other sources of V1-independent visually evoked excitation from as yet uncharacterized cortical or subcortical sources may have contributed to the recorded excitation. Some of the experiments illustrated in Extended Data Fig. 1b appear in previous publications and were also performed by one of the authors (PV-ChR2: all loose patch recordings are from Fig. 1b in ref. 32; VGAT-ChR2: 64 out of 138 units are from supplementary figures 7g, h in ref. 47).

**In vivo whole-cell recording.** Whole-cell recordings were made in V1 using the blind-patch technique<sup>48</sup> at a depth of 300–550 μm. This range of depth corresponds largely to the radial extent of L4 but may include some deep L2/3 and superficial L5 neurons. Patch pipettes (4–6 MΩ resistance) were pulled from borosilicate glass and filled with intracellular solution. Potassium-based intracellular solution (in mM: 135 K-gluconate, 8 NaCl, 10 HEPES, 4 Mg-ATP, 0.3 Na-GTP, 0.3 EGTA, pH 7.4) was used in 54 cells (52 recorded in both current clamp and voltage clamp, 2 recorded in voltage clamp only). 12 cells were recorded with caesium-based intracellular solution (in mM: 125 Cs-methanesulfonate, 8 NaCl, 10 HEPES, 4 Mg-ATP, 0.3 Na-GTP, 0.3 EGTA, 2 QX-314, pH 7.4) in voltage clamp only. Before the insertion of patch pipettes, a drop of 1.5% low-melting point agarose dissolved in ACSF (in mM: 140 NaCl, 5 KCl, 10 D-glucose, 10 HEPES, 2 CaCl<sub>2</sub>, 2 MgSO<sub>4</sub>, pH 7.4) was applied to the brain surface to reduce movement. Pipettes were rapidly inserted into the visual cortex while applying high positive pressure (2.5 psi). The pressure was reduced to 0.5 psi upon reaching a depth of 200–300 μm and the pipette was advanced in 2-μm steps while monitoring the pipette resistance. When a sudden increase in resistance was encountered, positive pressure was released and light suction was applied to achieve a gigaohm seal. Brief pulses of suction were applied to break the seal and achieve whole-cell configuration. Series resistance was 20–50 MΩ. Signals were amplified (Multiclamp 700B, Molecular Devices) and digitized at 10 kHz (DigiData, Molecular Devices) or 31.25 kHz (PCIe-6259, National Instruments). Membrane potential and spiking responses were recorded in current-clamp configuration. Excitatory currents were recorded in voltage-clamp configuration at –70 mV near the reversal potential of inhibition. Hence, the inhibitory currents generated by the optogenetic activation of inhibitory neurons for cortical silencing (see below) should not affect the recorded excitatory currents. In the case of improper space clamp, the inhibitory currents may affect the recorded excitation, but will do so similarly for all visual stimuli irrespective of direction. This is because, first, optogenetic photo-stimulation is the same, independent of the visual stimulus. Second, even if, on top of the optogenetic activation, some inhibitory neurons were to respond to the visual stimulus (for example, those inhibitory neurons that receive direct thalamic input), their visual response is likely to be the same irrespective of stimulus direction because inhibition recorded in layer 4 cortical neurons show no directional selectivity<sup>22</sup>.

**dLGN unit recording.** Before initiating V1 whole-cell recordings, a 4-shank (200 μm separation between shanks), 32 channel silicon probe (Buzsaki32, NeuroNexus) was inserted into the dLGN craniotomy (coordinates listed above) with the shanks distributed along the anterior/posterior axis. The probe was inserted at a 55° angle above horizontal in the coronal plane such that it advanced along the lateral to medial and dorsal to ventral directions. The probe was advanced in 5-μm steps until robust visual responses were observed in the multiunit activity (2.5–2.8 mm distance). The shanks typically spanned the anterior/posterior extent of the dLGN from 2–3 mm posterior of bregma. Retinotopy, as assessed with coarse receptive field mapping of multiunit activity, was consistent with a previous report<sup>37</sup>. Signals were amplified (Model 4000, AM Systems) and digitized at 31.25 kHz (PCIe-6259, National Instruments). For technical reasons, the top-most site on each shank was not recorded. The probe was allowed to settle for 30–60 min before collecting data. At the end of the experiment, the mouse was euthanized

under deep anaesthesia and the brain was fixed in 4% paraformaldehyde. The tissue was sectioned for post-hoc verification of the recording sites.

To improve the chances of recording from synaptically connected dLGN and V1 neurons, coarse receptive field maps of the dLGN multiunit activity were generated and the corresponding retinotopic region of V1 was identified by mapping the receptive field of the LFP signal at various locations in the V1 craniotomy. V1 LFP was recorded using a patch pipette filled with ACSF inserted to a depth of 200–400  $\mu\text{m}$ . The number of V1 locations that were sampled before finding the cortical location with receptive field overlap with the dLGN recording was approximately 1–5, although this was not precisely documented during the experiments. Subsequent V1 whole-cell recordings were targeted to this region of V1.

The UltraMegaSort<sup>49,50</sup> spike-sorting software was used to detect, cluster, and assign spike waveforms from dLGN recordings into single units as previously described<sup>51</sup>. Spike waveforms on the seven recorded sites of an individual shank were clustered using a *k*-means algorithm followed by manual assignment of clusters with distinct waveform profiles into single units.

**Identification of monosynaptic connections.** Out of 40 L4 whole-cell recordings in 24 mice and a total of 739 isolated thalamic units, 23 thalamocortical pairs (from 15 of the 24 mice) were considered monosynaptically connected according to the criteria described below. Those 23 connected pairs consisted of 17 L4 neurons and 23 thalamic units, because although for most L4 neurons we found only one presynaptic thalamic unit, for two L4 neurons we isolated two presynaptic units and in two L4 neurons we isolated three presynaptic units (Extended Data Fig. 6). In all connected pair analyses, each pair was treated independently without regard for whether or not the postsynaptic neuron was common to other connected pairs. As such, the postsynaptic responses of a cortical neuron may be represented several times if more than one pre-synaptic thalamic unit to that cortical neuron was identified (for example, the pairs in Fig. 3a).

Thalamocortical pairs were identified on the basis of two criteria illustrated in Extended Data Fig. 5. Criterion 1 sets a threshold for the thalamic unit spike-triggered average of the time derivative of the current recorded in L4 cortical neurons during cortical silencing and drifting grating presentation. Criterion 2 sets a threshold and a time window for the distribution of events detected in the time derivative of the L4 current (*I*) around the time of the spike in the thalamic unit. Both criteria have to be satisfied for the thalamic unit and the L4 cortical neuron to be considered a connected pair. We first calculated the time-derivative of the current recorded during cortical silencing in response to drifting gratings ( $dI/dt$ ). The thalamic unit spike-triggered average of  $dI/dt$  was computed and *z*-scored. If the *z*-score was less than a threshold of  $-5$  in the time window of 1–4 ms after the thalamic spike, the unit met criterion 1 and was considered a candidate presynaptic unit.

To ensure that the negative peak in the spike-triggered average of  $dI/dt$  was due to an actual increase in the probability of fast inward events occurring in the 1–4-ms window after the spike, we detected the occurrence of such events in the individual non-averaged  $dI/dt$  sweeps. Events were detected whenever  $dI/dt$  crossed a threshold of  $-1$  standard deviation. To calculate the standard deviation, we used all 50-ms intervals of  $dI/dt$  centred around each of the spikes of the thalamic unit. A perispike time histogram (PSPTH) was assembled from these event detections (0.2-ms bin size) and *z*-scored. Candidate presynaptic units with a *z*-score exceeding 3.5 in the time window 1–4 ms after the thalamic spike were considered monosynaptically connected to the cortical neuron.

To calculate  $dI/dt$ , the current (*I*), originally acquired at 31.25 kHz, was resampled at 10 kHz and differentiated in time using consecutive samples and smoothed (0.3-ms running average). Before *z*-scoring, we removed slow fluctuations from the spike-triggered average of  $dI/dt$  and the PSPTH by performing the following: We subtracted a smoothed version (3-ms running median) of the spike-triggered average of  $dI/dt$  and the PSPTH from the spike-triggered average of  $dI/dt$  and the PSPTH, effectively high-pass filtering them. The *z*-score was then calculated using the average and standard deviation of time points 0–5 ms before the thalamic spike.

**Properties of monosynaptic connections.** Latency of monosynaptic response was defined as the time of the peak in the *z*-scored PSPTH (see above) relative to the thalamic spike. Jitter was defined as the half-width at half-maximum of this peak (Extended Data Fig. 5).

The uEPSC was derived from the spike-triggered average of the current recorded in the L4 neuron during cortical silencing while presenting drifting gratings (Extended Data Fig. 5). The uEPSC may ride on top of a slower component of thalamic excitation generated by the response of other inputs to the visual stimulus. This component was estimated by shifting the trial number of thalamic unit responses by one trial relative to that of thalamic excitation for each direction of the drifting grating and computing a spike-triggered average on the basis of the trial-shifted data<sup>35</sup>. The trial-shifted spike-triggered average was subtracted from the uEPSC (shift-subtracted uEPSC) for amplitude and contribution (see ‘Thalamic unit contribution’ analyses. In the main figures, uEPSCs are shown

without shift-subtraction. Shift-subtracted uEPSCs are shown in Extended Data Fig. 6.

The amplitude of monosynaptic connections was defined as the most negative (inward) value of the shift-subtracted uEPSC from 0–3 ms after the onset of monosynaptic response (that is, the time of thalamic spike + latency for each pair; latency calculated using the PSPTH described above) minus the average value in a 0.2-ms window just before the onset. Amplitude of unitary excitatory postsynaptic potential (uEPSP) and spontaneous uEPSC (Extended Data Fig. 5) were determined in the same manner using the shift-subtracted spike-triggered average of  $V_m$  during drifting gratings under non-silencing conditions or the spike-triggered average of thalamic excitation during the 500 ms before visual stimulation during cortical silencing, respectively. Spontaneous uEPSCs were only characterized for pairs with at least 30 spontaneous spikes. One thalamic unit that passed the criteria for monosynaptic connection did not exhibit a clear spontaneous uEPSC in the simultaneously recorded cortical neuron despite firing a sufficient number of spontaneous spikes, and hence was not considered to be monosynaptically connected. In the four experiments in which several presynaptic thalamic neurons were recorded, the spike-sorting of monosynaptically connected thalamic units was verified using KiloSort software<sup>51</sup>. For one of these experiments, this yielded an additional thalamic unit which passed the criteria for monosynaptic connection and was included in the dataset.

**Drifting grating analysis.** Drifting grating responses were evaluated in a time window (the response window) from 0.3 s after stimulus onset to the end of the stimulus (2 s total, or 4 complete cycles). Spikes in current-clamp recordings were detected by identifying time points at which the membrane potential exceeded  $-15$  mV. The time of a spike was defined as the time of the peak depolarization in a 1.5-ms window after each threshold crossing. Before calculating the DSI of F1  $V_m$  (Fig. 1g), all spikes were removed from current-clamp recordings by replacing the time window from 1 ms before to 2 ms after the spike time with linear interpolation. In voltage-clamp recordings, the holding current in a 0.4-s window before the onset of visual stimulus was subtracted for each trial. This holding current was computed from the bottom 5th percentile of the distribution of current values in the 0.4-s window, which should include the periods with the least amount of spontaneous excitatory activity.

**Direction selectivity index.** The DSI for responses to two drifting grating stimuli moving in opposite directions was calculated as  $(\text{RespPref} - \text{RespNonPref}) / \text{RespPref}$ , where RespPref and RespNonPref are the size of the responses to the directions that gave the larger and smaller responses, respectively. This limits the range of DSI to between 0 and 1. A DSI of 0.3 means that the response in the preferred direction is 43% larger than in the non-preferred direction. For L4 neurons, the DSI was calculated for the spike rate, the F1 amplitude modulation of  $V_m$  (F1  $V_m$ ), the F1 amplitude modulation of thalamic excitation (F1Thal), the thalamic excitatory charge (QThal), and for the summation of the response to static gratings (F1Stat). For dLGN units, the DSI was calculated for the spike rate (F0) and the F1 modulation of spike rate.

The DSI of L4 neuron and dLGN unit spiking (Fig. 1g, left) was calculated from the average spike rate during the response window (see definition of response window above). F1 amplitude was derived from the amplitude of a sinusoidal fit to the cycle average of drifting grating responses. The period of the cycle average was 0.5 s, matching the temporal frequency (2 Hz) of the drifting grating. QThal was calculated from the integral of drifting grating excitatory current over the response window. To determine the DSI of F1Thal after arithmetic equalization of QThal generated by the two opposite stimulus directions (Extended Data Fig. 2), we multiplicatively scaled the thalamic excitatory current such that QThal was identical for both directions.

When the DSI was compared for two different response parameters within the same cell for example, DSI F1Thal versus DSI F1  $V_m$  (Fig. 1g middle), the DSI was defined relative to the preferred direction of one of the parameters, the reference parameter. If the preferred direction of the two parameters were different, the DSI of the non-reference parameter was multiplied by  $-1$ . For example, in Fig. 1g middle, the reference parameter was  $V_m$ . Hence, the DSI of the reference parameter is always positive (range 0 to +1) but the DSI of the non-reference parameter can be negative or positive (range  $-1$  to +1) with negative and positive values indicating that the two parameters prefer opposite or the same direction, respectively. The absolute value of the DSI indicates the degree of selectivity.

**Static grating analysis.** Static grating analysis was restricted to responses collected under cortical silencing conditions. The thalamic EPSC in response to each spatial phase of the static grating were averaged together. The average value of the current from 0 to 24 ms after stimulus onset was subtracted. This time window was before any observable visual response.

Summation of static grating responses to simulate drifting gratings was performed by taking the thalamic EPSCs evoked by each of the 16 phases of the static gratings and staggering them in time (by 31.25 ms, that is, by 1/16 of the period of the drifting grating) to match the temporal sequence at which each of the 16

individual spatial phases would occur during a drifting grating with temporal frequency of 2 Hz (Fig. 2c, d). We also staggered the gratings by shorter intervals to mimic higher temporal frequencies (Extended Data Fig. 3). We compared the algebraic sum in which EPSCs were ordered according to the spatial phase sequence simulating the motion of the grating in one direction against the sum simulating motion in the opposite direction. The DSI and F1 amplitudes of the resulting algebraic summations (DSI F1Stat) were analysed as described above. The spatial phases were ordered such that the relationship between the sequence of spatial phases and the preferred direction of F1Thal was the same for each neuron. Specifically, the preferred direction of F1Thal is in the direction of increasing spatial phase, that is, the upward direction for static grating heat maps (Figs. 2f, 4a, b).

Early and late thalamic excitation were defined as the excitatory charge from 30 to 110 ms (early EPSC) and from 110 to 230 ms (late EPSC) after static grating onset, respectively. For the plots of the phase-dependence of early and late EPSCs, data were normalized by the peak value across all spatial phases. The preferred spatial phases of early and late thalamic excitation were calculated from the vector average across all spatial phases. Thus, the preferred spatial phase may fall in between one of the 16 presented phases. Accordingly, for the population average of the phase-dependence of early and late EPSCs (Fig. 2e, right), we upsampled the 16 data points of each cell to 96 via linear interpolation. This enabled us to align the cells by the preferred spatial phase of the early EPSC. The preferred phase of the early EPSC was set at 180°.

For spatiotemporal receptive field analysis, the preferred spatial phase was calculated for each of five contiguous time bins (each of 40-ms duration, starting 30 ms after stimulus onset). A linear fit to these five preferred spatial phase values was performed to calculate the spatiotemporal slope. For population averages of static grating heat maps (Fig. 2f, right; Fig. 4b), the static grating responses of each cortical neuron were normalized by the peak response across all spatial phases, upsampled in the spatial phase dimension as above and shifted along the spatial phase axis so that the preferred spatial phase of the earliest time bin (30–70 ms) occurred at a spatial phase of 180°.

**dLGN unit analysis.** *Static grating.* For each of the 739 thalamic units, peristimulus time histograms (PSTH; 10-ms binning, upsampled to 10 kHz, smoothed by 20-ms running average) of the spiking were constructed for responses to each spatial phase of the static grating under cortical silencing conditions. For analysis of the time course of thalamic unit responses across different spatial phases (Extended Data Fig. 4), the 177 units were selected on the basis of the following two criteria: the average firing rate across the stimulus duration (250 ms) exceeded 16 Hz in response to the preferred spatial phase of the unit and in response to spatial phases 45° offset from the preferred; and the peak firing rate exceeded 25 Hz in response to the preferred spatial phase of the unit and in response to spatial phases 45° offset from the preferred.

*Drifting grating.* To calculate the DSI of thalamic units in response to drifting gratings (Extended Data Figs. 4, 7) PSTHs of the spiking response to drifting gratings were constructed for each thalamic unit. For the DSI of the F1 modulation, the F1 amplitude of the thalamic unit response was derived from the amplitude of a sinusoidal fit to the cycle average of the drifting grating PSTH. For the DSI of the average spiking response (F0) we used the average firing rate across the duration of the stimulus.

**Analysis of the time course of EPSC and PSTH in response to static gratings.** Out of 23 pairs, 3 of the pre-synaptic thalamic units had poorly defined spatiotemporal receptive fields in response to static gratings (units with asterisks in Extended Data Fig. 6) and were not included in the analysis in Fig. 3.

The duration of the EPSC for each spatial phase was defined as the time point after stimulus onset at which the integral of the response reached 90% of its maximum value. To compare EPSC and PSTH time courses, only spatial phases in which both the PSTH and the EPSC were at least 10% of that elicited by the spatial phase that gave the largest response were included in Fig. 3e, f. Each spatial phase response was peak-normalized, and the Pearson's correlation between the EPSC and the PSTH was computed and then averaged across all spatial phase responses. To test statistical significance of the average pairwise Pearson's correlation between EPSC and PSTH spatial phase responses, each EPSC response was reassigned to a random PSTH response. For each shuffle, the average pairwise correlation was calculated, and this was repeated for 10,000 shuffles. The average pairwise correlation of the real data was compared to the distribution of shuffled average pairwise correlations. None of the shuffled average pairwise correlations exceeded that of the real data. Use of the Spearman correlation produced similar results.

**Compound neuron.** For static gratings (Fig. 4b), the heat maps of the PSTH and EPSC were shifted along the spatial phase axis so that the preferred spatial phase of the earliest time bin (30–70 ms) of the EPSC occurred at a spatial phase of

180°. The sequence of spatial phases was ordered so that the preferred direction of thalamic excitation (F1Thal) in response to drifting gratings was in the upward direction, that is, the direction of increasing spatial phase. Heat maps were normalized by the peak response across all spatial phases, upsampled in the spatial-phase dimension as above, and averaged across the eight pairs in which DSI F1Thal was greater than 0.3.

For drifting gratings, the cycle-averaged responses to drifting gratings of the preferred and non-preferred direction of the EPSC (F1Thal) were combined across pairs. Before combining, the EPSC and PSTH cycle-average for each pair was shifted in time by the same amount, so that the F1 peak of the EPSC occurred at 250 ms. Responses were normalized using the direction that produced the largest amplitude. Only pairs in which DSI F1Thal was greater than 0.3 were included.

**Estimation of the number of thalamic units contributing to visually evoked response in an L4 neuron.** The excitatory current contributed by the thalamic neuron (Fig. 6) was calculated for each connected pair by convolving its shift-subtracted uEPSC from 0–15 ms after the thalamic spike with its spike train during drifting gratings under cortical silencing conditions and averaging across trials. The shift-subtracted uEPSC was truncated to the time point at which it returned to the baseline if this occurred before 15 ms. The contributed charge was the integral of the trial-averaged convolution across the response window. Dividing the contributed charge by the thalamic excitatory charge evoked by the same drifting grating stimulus resulted in the unitary contribution of the thalamic neuron. Estimates of thalamocortical convergence from the distribution of unitary contributions are analysed in Extended Data Fig. 8.

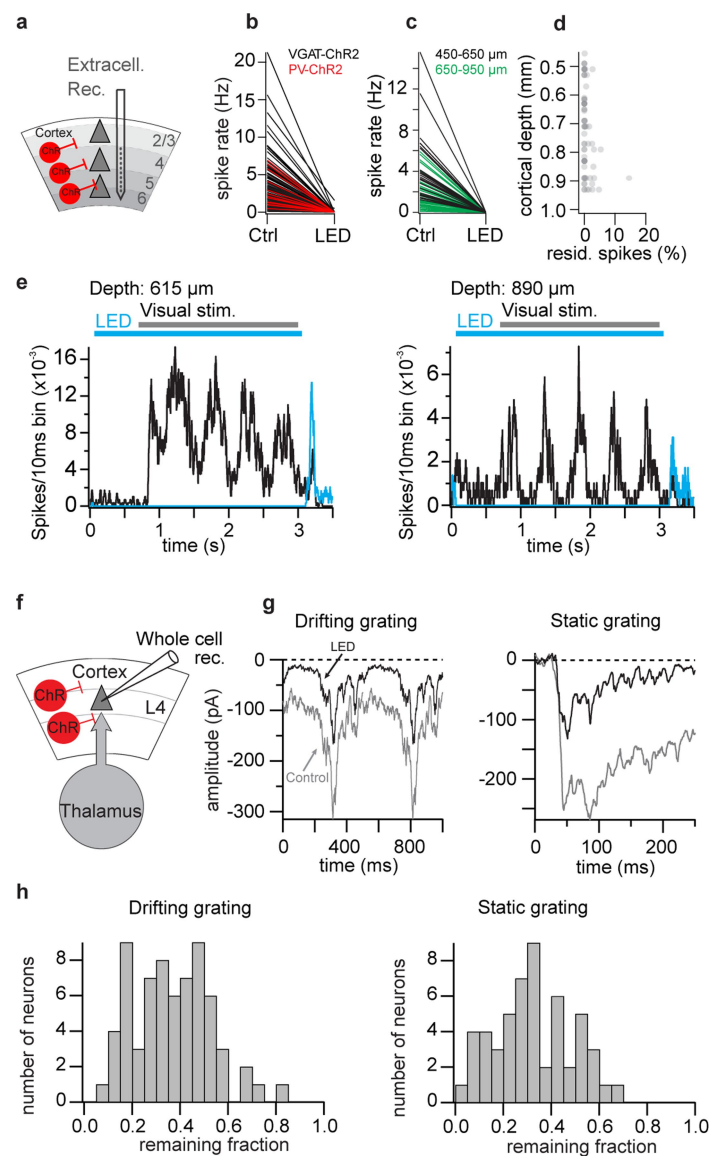
**Model.** The model (Fig. 5) consisted of a transient thalamic neuron and a sustained thalamic neuron converging on the same cortical neuron. The time courses of EPSCs generated by the transient and sustained thalamic neurons in response to static gratings were modelled with an instantaneous rise followed by a single exponential decay. The decay time constant of the sustained input was longer than that of the transient input. For each input, the time course of the response to static gratings of different spatial phases was fixed, but the amplitude was modulated sinusoidally across spatial phase with a period of one cycle. The response to 32 evenly spaced spatial phases covering one cycle were generated. The compound thalamic excitation from both inputs was derived from the linear summation of the transient input and sustained input EPSCs. The compound thalamic excitation during drifting gratings was derived from static grating responses (summation after staggering in time to mimic drifting gratings in either direction) and the F1 modulation and its DSI were calculated as described above. We systematically varied the preferred spatial phase of transient and sustained inputs to see how their relative phase shift affected direction selectivity. The effect of different sustained input decay time constants was also examined.

**Statistical analysis.** Statistical analyses were performed in IgorPro. No statistical methods were used to predetermine sample sizes, but our sample sizes were similar to those reported in previous publications in the field. All data are presented as mean  $\pm$  s.d. unless otherwise specified. Normality of the data were not tested and non-parametric two-sided Wilcoxon rank-sum or Wilcoxon signed-rank tests were used for unpaired or paired tests, respectively, unless otherwise specified. The fraction of cells with matching direction preference was compared to a chance value of 0.5 using a two-tailed binomial test. Experiments and analysis were not blinded.

**Reporting summary.** Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

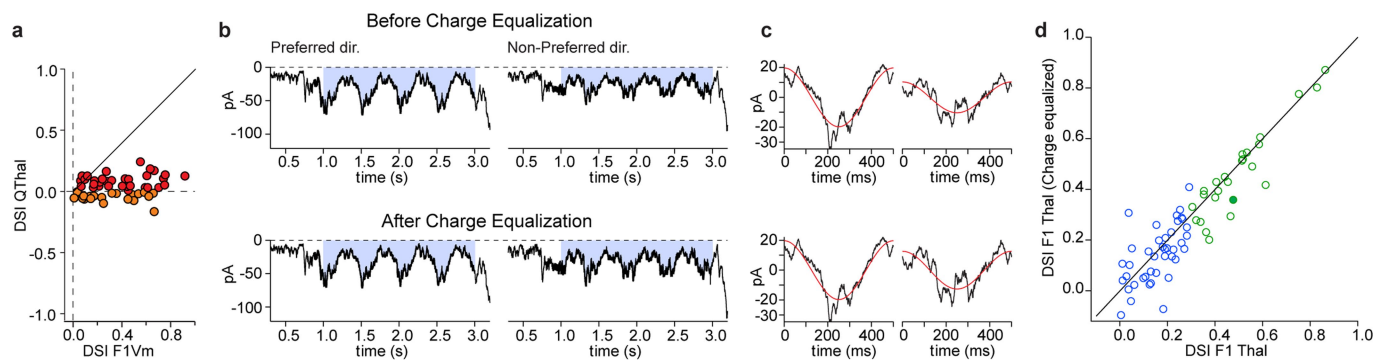
**Code and data availability.** Custom code used and datasets generated and/or analysed during the current study are available from the corresponding author upon reasonable request.

45. Hippenmeyer, S. et al. A developmental switch in the response of DRG neurons to ETS transcription factor signaling. *PLoS Biol.* **3**, e159 (2005).
46. Zhao, S. et al. Cell type-specific channelrhodopsin-2 transgenic mice for optogenetic dissection of neural circuitry function. *Nat. Methods* **8**, 745–752 (2011).
47. Saleem, A. B. et al. Subcortical source and modulation of the narrowband gamma oscillation in mouse visual cortex. *Neuron* **93**, 315–322 (2017).
48. Margrie, T. W., Brecht, M. & Sakmann, B. In vivo, low-resistance, whole-cell recordings from neurons in the anaesthetized and awake mammalian brain. *Pflügers Arch.* **444**, 491–498 (2002).
49. Fee, M. S., Mitra, P. P. & Kleinfeld, D. Automatic sorting of multiple unit neuronal signals in the presence of anisotropic and non-Gaussian variability. *J. Neurosci. Methods* **69**, 175–188 (1996).
50. Hill, D. N., Mehta, S. B. & Kleinfeld, D. Quality metrics to accompany spike sorting of extracellular signals. *J. Neurosci.* **31**, 8699–8705 (2011).
51. Olsen, S. R., Bortone, D. S., Adesnik, H. & Scanziani, M. Gain control by layer six in cortical circuits of vision. *Nature* **483**, 47–52 (2012).



**Extended Data Fig. 1 | Cortical silencing** **a**, Experimental configuration. Channelrhodopsin-2 (ChR2) is expressed in cortical inhibitory neurons to suppress neuronal activity upon illumination with a blue LED while performing extracellular recordings. **b**, Visually evoked activity (full-field drifting gratings) from units isolated throughout the cortical depth is suppressed upon LED illumination. Black lines, ChR2 was expressed in all GABAergic neurons (GABA,  $\gamma$ -aminobutyric acid; VGAT-ChR2 mice; 138 units recorded with silicon probes;  $98.9 \pm 2.7\%$  silencing in 25 units recorded above  $500 \mu\text{m}$  depth;  $99.4 \pm 2\%$  silencing in 113 units recorded below  $500 \mu\text{m}$ ; 5 mice). Red lines, ChR2 was conditionally expressed in parvalbumin (PV)-expressing neurons through viral injection into the visual cortex of the PV-Cre mouse line (13 loose patch recordings in layer 4; 2 mice; 100% silencing). **c**, As in **b** but specifically for units recorded between  $450\text{--}650 \mu\text{m}$  depth (black;  $99.8 \pm 0.6\%$  silencing;  $n = 26$ ; 3 mice) and  $650\text{--}950 \mu\text{m}$  depth (green; putative layer 6;  $99.8 \pm 2.4\%$  silencing;  $n = 48$ ; 3 mice). These units are a subset of the units from VGAT-ChR2 mice illustrated in **b** for which the recording depth could be estimated. **d**, Percentage of visually evoked spikes remaining during LED illumination

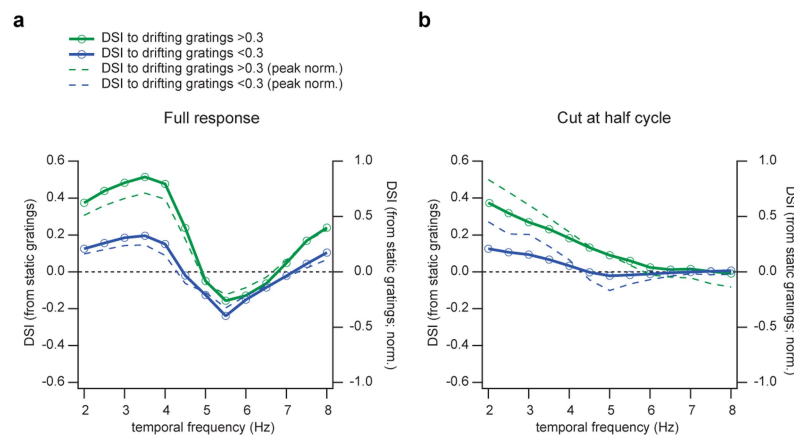
across cortical depths deeper than  $450 \mu\text{m}$ . The units are the same as in **c**. **e**, Peristimulus time histogram of two units located at  $615 \mu\text{m}$  (left) and  $890 \mu\text{m}$  (right) depth in response to drifting gratings under control conditions (black) and during cortical silencing (blue). The duration of the visual stimulus and of the LED illumination is illustrated by the horizontal bars. **f**, Experimental configuration. As in **a** but whole-cell recordings from layer 4 neurons instead of extracellular recordings. **g**, Whole-cell voltage-clamp recording ( $V_{\text{holding}} = -70 \text{ mV}$ ) of a layer 4 neuron (same neuron as in Fig. 2a). Response to drifting gratings (left; two identical cycle averages are shown for clarity) and static gratings (right; average of 10 traces). The grey indicates control conditions; the black trace was recorded during LED illumination to isolate thalamic excitation. **h**, Distribution of remaining excitatory charge upon LED illumination for drifting gratings (66 recordings as in **g**, left) and static gratings presented at the preferred spatial phase (53 recordings as in **g**, right). The visually evoked excitation was reduced by about 65% (drifting grating,  $63 \pm 16\%$ ; static grating,  $68 \pm 16\%$ ).



### Extended Data Fig. 2 | The contribution of the directional preference of the thalamic charge to the direction selectivity index of F1Thal.

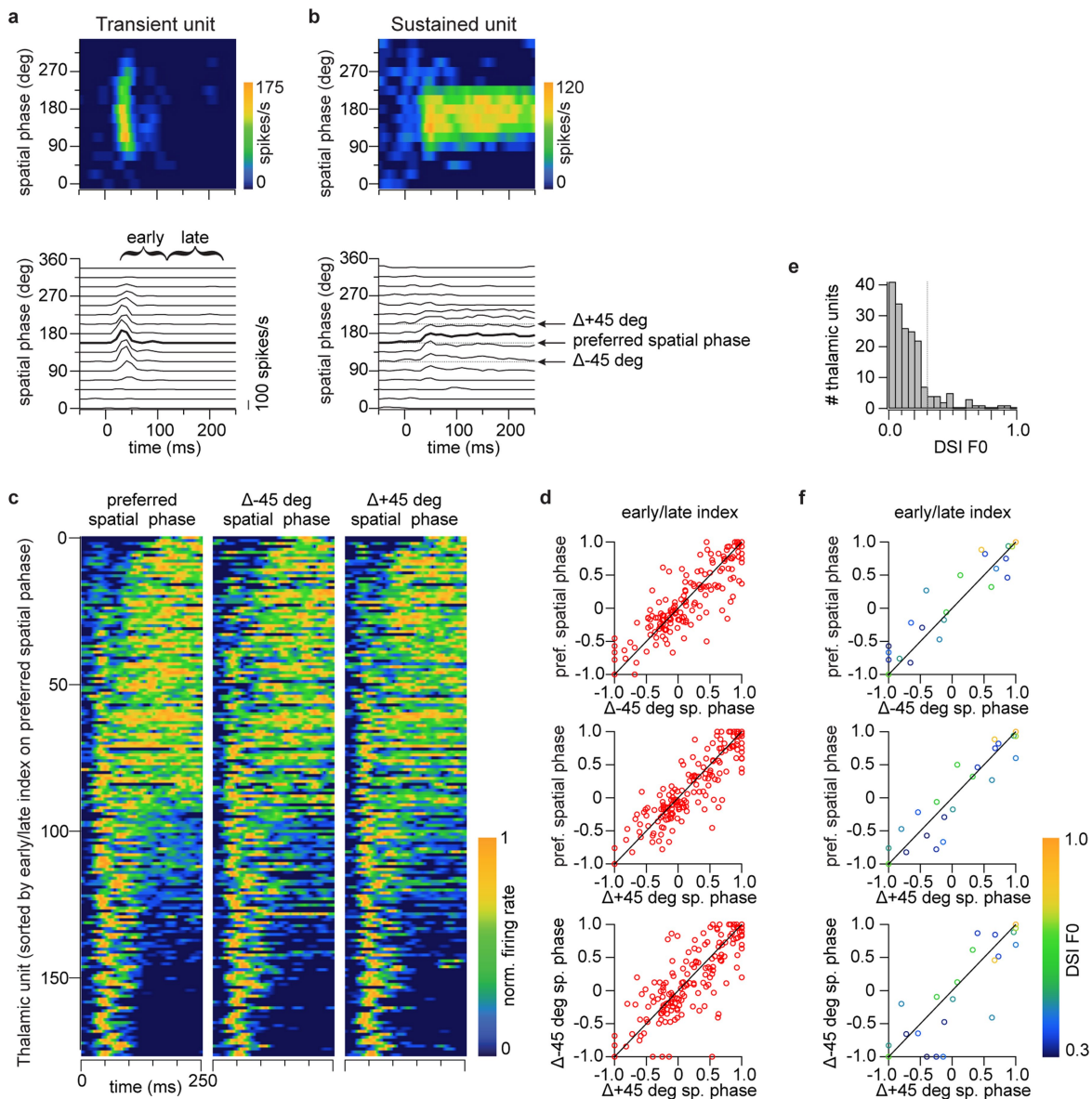
**a.** The same plot as in Fig. 1g, right. The data points for which QThal and F1Thal prefer the same direction (positive values on the y axis) are shown in red; data points for which QThal and F1Thal prefer opposite directions (negative values on the y axis) are orange. Note that the average absolute value of DSI QThal is slightly but significantly larger for the red data points (red points:  $0.087 \pm 0.053$ ;  $n = 32$ ; orange points:  $0.048 \pm 0.037$ ;  $n = 20$ ;  $P = 0.003$ ;  $t$ -test), indicating a slight bias of QThal in the preferred direction. To determine the effect of this slight bias in thalamic charge on DSI F1Thal, we have equalized the thalamic charge evoked by gratings drifting in both directions (see **b–d**). **b.** Example recording from a cortical neuron for which the charge of thalamic excitation is larger in the preferred versus the non-preferred direction of F1Thal. Top, thalamic excitation as recorded (non-equalized) in response to a grating drifting in

the preferred (left) and non-preferred (right) direction. Bottom, same as top but after scaling the response to the non-preferred direction such that the charge is the same in either direction (charge equalization). **c.** Cycle average of thalamic excitation with superimposed sinusoidal fit (red). Top, as recorded; bottom, after charge equalization. After charge equalization the direction preference is maintained but, for this particular example, the DSI is reduced (see filled data point in **d**). **d.** Scatter plot for all recordings (green points indicate DSI F1Thal > 0.3; blue points indicate DSI F1Thal < 0.3). The filled data point is the example above. The equalization leads to only a small change in DSI F1Thal (all points:  $0.28 \pm 0.20$  before and  $0.26 \pm 0.21$  after equalization;  $P = 0.034$ ; paired  $t$ -test;  $n = 66$ ; green points:  $0.49 \pm 0.15$  before and  $0.46 \pm 0.17$  after equalization;  $P = 0.022$ ; paired  $t$ -test;  $n = 25$ ; blue points:  $0.16 \pm 0.09$  before and  $0.14 \pm 0.11$  after equalization;  $P = 0.300$ ; paired  $t$ -test;  $n = 41$ ).



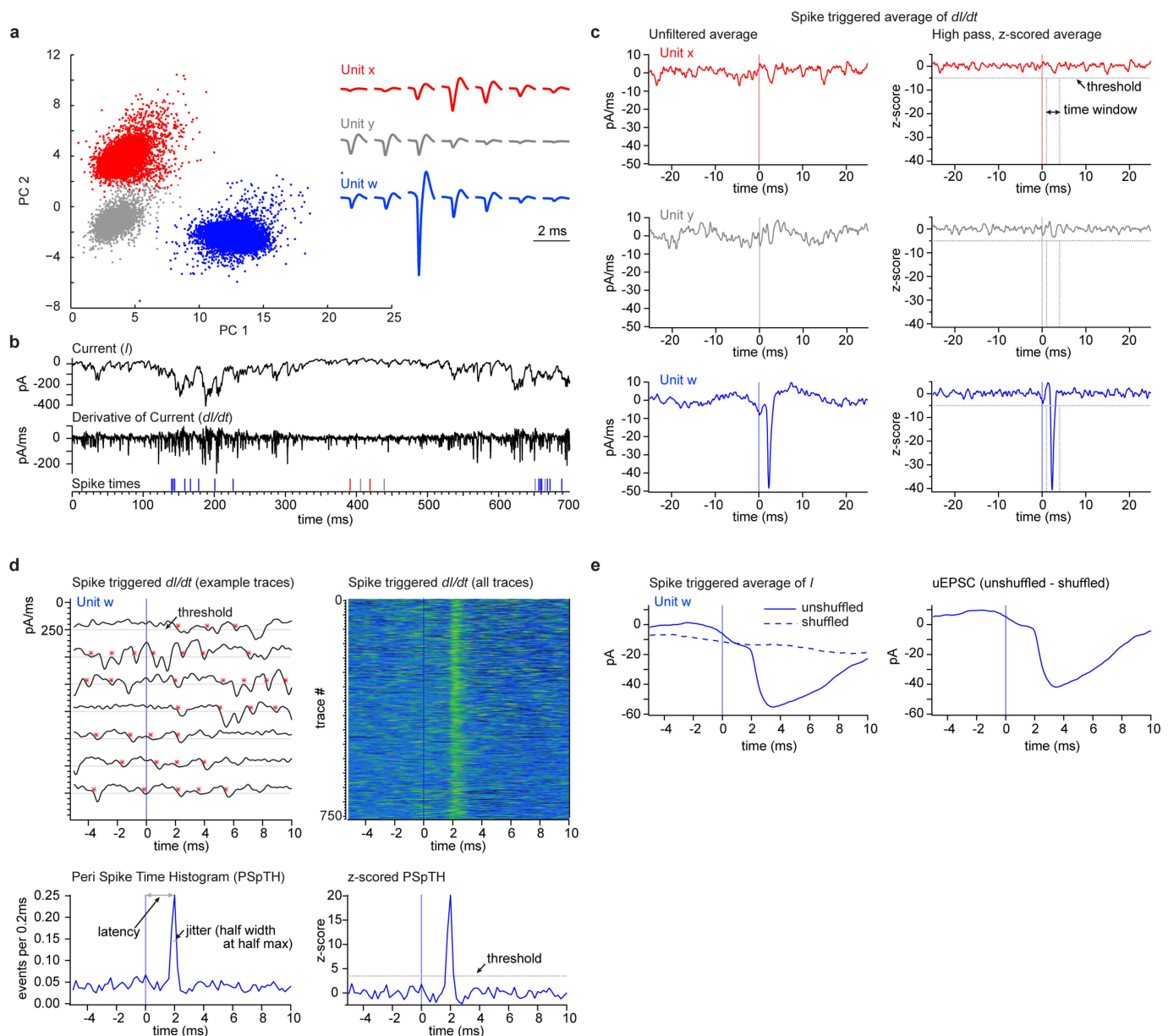
**Extended Data Fig. 3 | Predicting the DSI for various temporal frequencies from the response to static gratings.** The amplitude of the F1 modulation was determined from the algebraic sum of the thalamic EPSCs evoked by each of the 16 phases of the static grating, as in Fig. 2c, d. The thalamic EPSCs were staggered in time to mimic different temporal frequencies of a drifting grating (for example, at 4 Hz a cycle lasts 250 ms and hence the response to each one of the phases is staggered by 15.6 ms (250/16 ms) relative to the preceding one). The DSI was computed by comparing the F1 modulation of the sum in which EPSCs were ordered according to the spatial phase sequence simulating the motion of the grating in one direction against the sum simulating motion in the opposite direction. Green and blue traces, average of all cells for which DSI F1Thal

to drifting gratings was larger ( $n = 18$ ) or smaller ( $n = 28$ ) than 0.3, respectively. For the dotted traces, the computed DSI was normalized to the peak for each cell; right ordinate. **a**, The full 250-ms response to static gratings was used to compute the DSI at each temporal frequency. Note the reversal of direction preference at higher temporal frequencies. **b**, Only the initial  $x$  milliseconds of the response to static gratings were used to compute the DSI,  $x$  being the half period of the temporal frequency to be computed (for example, for 4 Hz,  $x = 125$  ms). The rationale for this approach is that the interactions between excitatory inputs that are relevant for the emergence of direction selectivity probably occur within a half cycle.



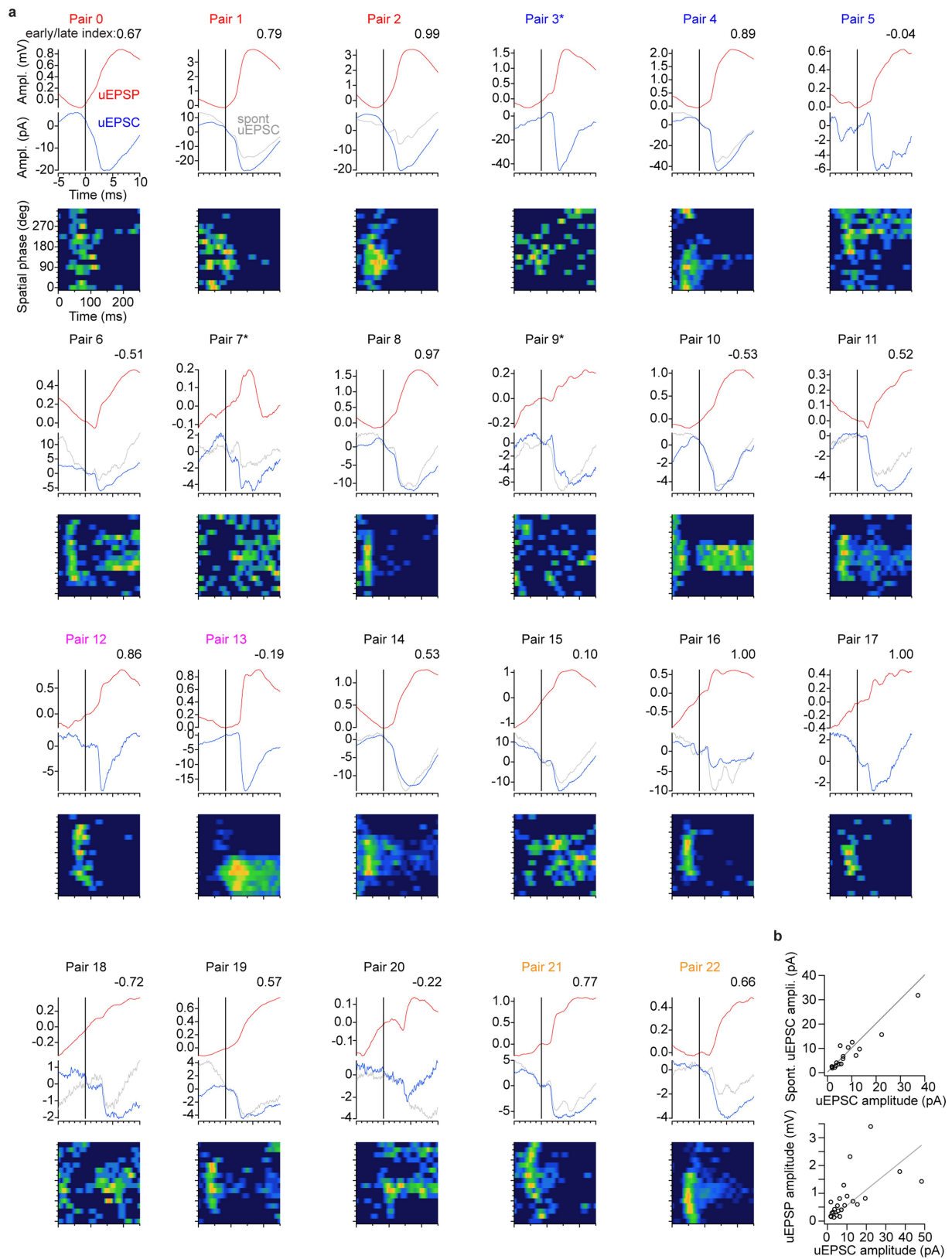
**Extended Data Fig. 4 | The time course of the activity of thalamic units recorded in dLGN in response to static gratings is similar across spatial phases.** **a**, Example thalamic unit with transient response to static gratings during cortical silencing. Top, spatiotemporal receptive field. Bottom, PSTHs in response to each phase of the static grating used to construct the spatiotemporal receptive field illustrated above. The PSTH at the preferred phase is highlighted by a thicker trace. The preferred phase is defined as the phase closest to the vector average of the response at each phase. The brackets show the time windows over which the early (30–110 ms) and late (110–230 ms) firing rates were averaged ( $R_e$  and  $R_l$ , respectively) to compute the early/late index  $[(R_e - R_l)/(R_e + R_l)]$ . This unit has an early/late index of 1 for static gratings presented at the preferred phase and of 0.88 and 1 for gratings presented at phases of  $\pm 45^\circ$  from the preferred phase. **b**, As in **a** but for an example thalamic unit with sustained response. The arrows illustrate the preferred spatial phase and the phases separated by  $\pm 45^\circ$ . This unit has an early/late index of  $-0.3$  for static gratings presented at the preferred phase and of  $-0.08$  and  $0.22$  for gratings presented at phases of  $\pm 45^\circ$  from the preferred phase. **c**, Heat maps of responses to static gratings for 177 thalamic units (24 mice) during

cortical silencing. Left, each row is the amplitude of the PSTH of one of the thalamic units in response to the preferred phase of the static grating. The units are ordered according to their early/late index in response to the preferred phase. Middle, same as left but in response to a static grating, the phase of which is  $45^\circ$  below the preferred phase. The order of the units has not been changed; that is, it is the same as on the left. Right, same as left but in response to a static grating with a phase  $45^\circ$  above the preferred phase. The order is the same as on the left. Note that transient and sustained units maintain their characteristic firing dynamics even in response to static gratings presented at phases of  $\pm 45^\circ$  from the preferred phase. **d**, Scatter plots of the early/late index computed in response to static gratings presented at the preferred phase and at phases of  $\pm 45^\circ$  from the preferred phase. Note that in all plots the data are close to the unity line. **e**, Distribution of direction selectivity indexes of the firing rates (DSI F0) of the 177 thalamic units in **d**. The vertical dotted line is DSI F0 = 0.3. In 22 units DSI F0 was greater than 0.3. **f**, As in **d** but specifically for those thalamic units with DSI F0 values greater than 0.3. Colours indicate the value of DSI F0, according to the colour scale on the right.



**Extended Data Fig. 5 | Criteria for identifying monosynaptically connected thalamocortical pairs.** Thalamocortical pairs were identified on the basis of two criteria: Criterion 1 (illustrated in **c**) sets a threshold for the thalamic unit spike-triggered average of the time derivative of the current recorded in L4 cortical neurons. Criterion 2 (illustrated in **d**) sets a threshold and a time window for the distribution of events detected in the time derivative of the L4 current around the time of the spike in the thalamic unit. Both criteria have to be satisfied for the thalamic unit and the L4 cortical neuron to be considered as a connected pair. **a**, Isolation of thalamic units in the dLGN. Left, first two principal components illustrating three separable clusters attributed to three independent thalamic units (units x, y and w in red, grey and blue, respectively). Right, electrophysiological recording illustrating the average spike shape recorded from seven electrodes for the three thalamic units. **b**, Differentiation of the current recorded in L4 neurons, the same experiment as in **a**. Top trace, the current recorded in the whole-cell configuration from an L4 cortical neuron ( $V_{\text{holding}} = -70$  mV) during the presentation of a drifting grating (single trial). Middle trace, the temporal derivative of the above current ( $dI/dt$ ). Lower, the times at which each one of the three thalamic units from **a** (x, red; y, grey; w, blue) fired during the same trial. **c**, Criterion 1. Left, spike triggered average of  $dI/dt$  of the current recorded in the L4 neuron for the three thalamic units illustrated in **a**. Time 0 denotes the time of the spike. Right, same spike-triggered averages shown on the left after high-pass filtering and z-scoring

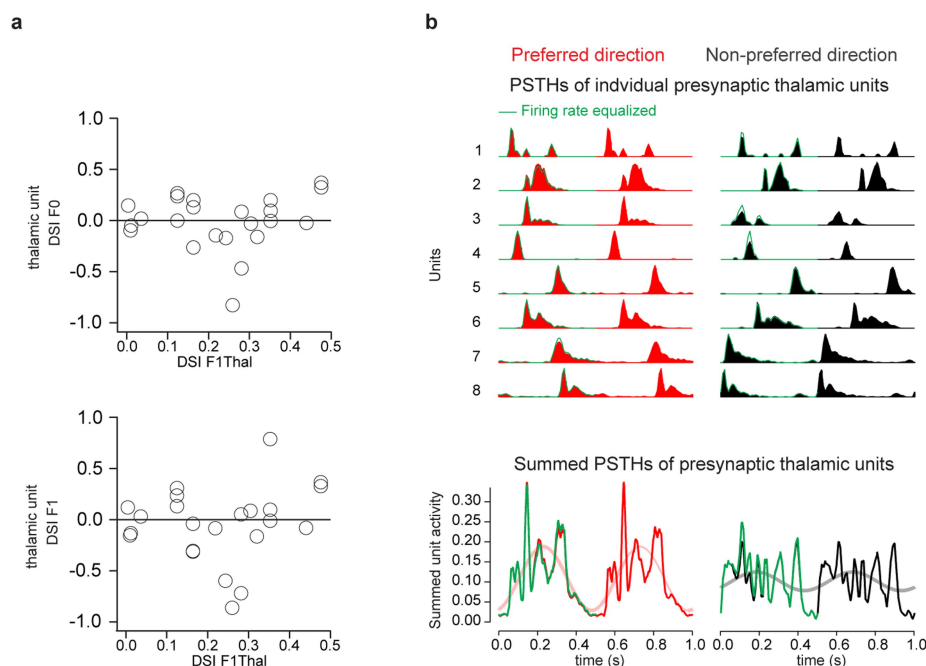
(see Methods). Note that only unit w (blue) crosses the 5z threshold. **d**, Criterion 2. Top left, seven individual time derivatives of currents ( $dI/dt$ ) recorded in the L4 neuron (same as in **b**) aligned relative to seven spikes recorded in unit w (time 0 denotes the time of the spike). Each asterisk shows an event crossing the threshold of  $-36 \text{ pA ms}^{-1}$ . Top right, same as left but represented as a heat map of the amplitude of  $dI/dt$  for 761 traces (the heat map colour scale ranges from  $+50 \text{ pA ms}^{-1}$  to  $-200 \text{ pA ms}^{-1}$ ). This heat map clearly illustrates an increase in event probability around 2 ms after the spike in unit w. Bottom left, the PSpTH for the events detected in the 761 traces illustrated above. The peak of the PSpTH is used to determine the latency (that is, the time interval between the spike recorded in the thalamic unit and the occurrence of a postsynaptic response detected in the L4 cortical neuron). The half width at half maximum is used to determine the jitter of that response (in this example the latency is 2 ms and the jitter is 188 microseconds). Bottom right, same as left but z-scored. The PSpTH must cross a threshold of 3.5z within 1–4 ms after the spike in the thalamic unit for the thalamic unit to be considered as synaptically connected to the L4 neuron. **e**, Left, unit w spike-triggered average of the response recorded in the same L4 neuron as in **a**. The continuous blue line represents unshuffled trials, the dotted line represents shuffled trials (see Methods). Right, the difference between the shuffled and unshuffled trials is used to isolate the uEPSC between unit w and the recorded L4 neuron.



Extended Data Fig. 6 | See next page for caption.

**Extended Data Fig. 6 | Unitary EPSCs, EPSPs and spatiotemporal receptive fields of presynaptic thalamic units. a,** Each panel illustrates one of the 23 paired recordings. The upper part of the panel shows the shift-subtracted uEPSP (red; Methods) and the shift-subtracted uEPSC (blue; Methods) recorded during visual stimulation. The spatiotemporal receptive field (heat map) of the presynaptic thalamic unit, obtained from the response to static gratings, is shown on the bottom. For some pairs the uEPSC recorded during spontaneous activity (spont uEPSC; grey; Methods) is also shown. uEPSCs are recorded during cortical silencing; uEPSPs are recorded under control conditions. The vertical line at time 0 marks the time of the peak of the extracellularly recorded action potential in the presynaptic thalamic unit. Pair numbers of the same colour were recorded in the same postsynaptic layer 4 cortical neuron. The heat map shows the spatiotemporal receptive field of the thalamic unit in response to static gratings. Each spatiotemporal receptive field is centred ( $157.5^\circ$ ) on the preferred spatial phase (defined as the phase that produced the most spikes) of its unit except for converging pairs which are aligned to the average preferred phase of the converging units. The response of pairs

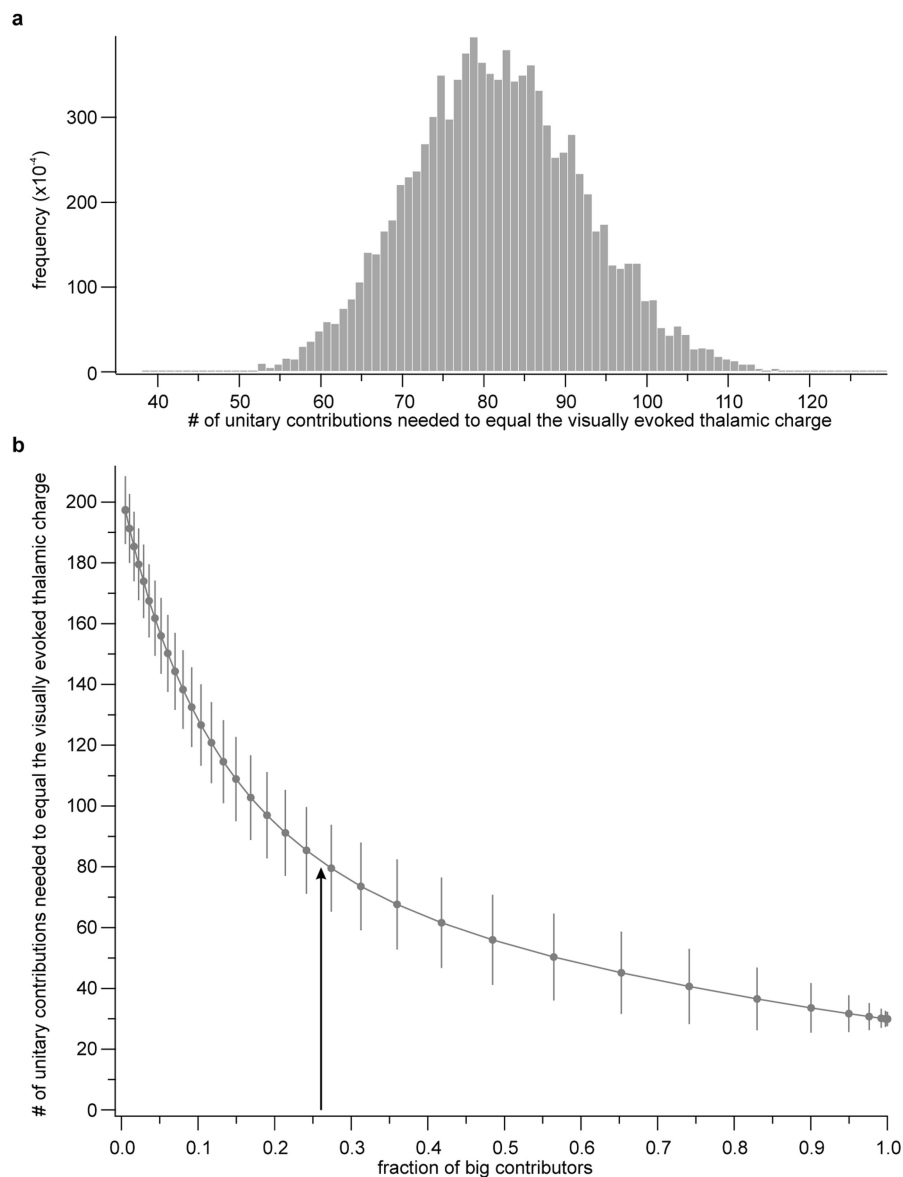
marked by an asterisk were not included in the analysis of static gratings because of their poorly defined spatiotemporal receptive field. The early/late index (see Extended Data Fig. 4) for the preferred phase of the presynaptic thalamic unit is given for each pair on the top right in black except for pairs with an asterisk. uEPSCs (blue) are the average of 49–970 spike-triggered traces. uEPSPs (red) are the average of 101–1,496 spike-triggered traces. Spontaneous uEPSCs (grey) are the average of 30–412 spike-triggered traces. Units of pairs 10, 11, 12, 13, 14, 18, 21 and 22 are the eight presynaptic units to the compound neuron in Fig. 4 and correspond to the units numbered in Fig. 4 as 6, 7, 1, 2, 8, 5, 4 and 3, respectively. Pair 4 corresponds to the pair shown in Fig. 6 and in Extended Data Fig. 5. **b,** Top, correlation between the amplitude of the visually evoked uEPSC (blue in **a**) and the spontaneous uEPSC (grey in **a**) for those pairs in which both could be recorded ( $r = 0.95$ ;  $P = 5.9 \times 10^{-9}$ ;  $n = 17$ ). The diagonal line indicates unity. Bottom, correlation between the amplitude of the visually evoked uEPSC (blue in **a**) and the uEPSP (red in **a**) for all pairs ( $r = 0.59$ ;  $P = 0.0028$ ;  $n = 23$ ). The diagonal line is a linear fit to the data with a slope of  $0.056 \text{ mV pA}^{-1}$ .



**Extended Data Fig. 7 | The direction selectivity of presynaptic thalamic units does not contribute to the direction selectivity of L4 neurons.**

**a**, Top, the DSI of the average firing rate of each presynaptic thalamic unit (DSI F0) is plotted against DSI F1Thal of the postsynaptic L4 cortical neuron. Bottom, the DSI of the F1 modulation of the firing rate of each presynaptic thalamic unit (DSI F1) is plotted against DSI F1Thal. Note the absence of correlation between DSI F1Thal and DSI F0 ( $r = 0.09$ ;  $P = 0.68$ ) and DSI F1 ( $r = 0.16$ ,  $P = 0.46$ ) of the thalamic unit. Also, note that in only about half of the pairs (12 out of 23 in the top graph; 11 out of 23 in the bottom graph) the presynaptic thalamic unit and the postsynaptic cortical neuron share the same preferred direction, as expected by chance ( $P = 1.00$  for both, binomial test). As such, the DSI of thalamic units does not predict DSI F1Thal of the cortical neuron. **b**, Top, the PSTHs of each

of the eight units contributing to the compound neuron (from Fig. 4). The units were temporally aligned relative to each other using the phase of the F1 modulation of thalamic excitation recorded in their postsynaptic L4 target neurons in response to gratings drifting in the preferred (red) and non-preferred direction (black). Two identical cycles are shown for clarity. The equalized PSTHs (that is, the PSTHs that were scaled such that the firing rate of the thalamic unit is the same in either direction) are shown in green. Only the first cycle is equalized to facilitate comparison. Bottom, summed PSTHs of the eight presynaptic thalamic units (pink and grey lines are sinusoidal fits; from Fig. 4). The green traces are the summed activity of the equalized PSTHs. Note the similarity between the control and the equalized summed activity.



**Extended Data Fig. 8 | Number of thalamic neurons contributing to the visually evoked response of an L4 neuron.** **a**, Distribution of the number of unitary contributions of thalamic neurons necessary to equal the total thalamic charge recorded in an L4 cortical neuron in response to a grating drifting in the preferred direction during cortical silencing. The distribution is obtained by randomly sampling with replacement the individual contributions from each of the 23 pairs, 10,000 times. In each iteration, unitary contributions were sampled until their sum reached 100%. To compute the unitary contribution of a thalamic unit, we first convolved the spike train of the unit in response to a drifting grating with the uEPSC that that unit evoked in the postsynaptic L4 cortical neuron during cortical silencing. We then integrated the resulting current in time and normalized the obtained charge by the total charge recorded in the postsynaptic cortical neuron in response to the drifting grating, also during cortical silencing. Unitary contributions are expressed as a percentage of the total charge. On average,  $80.9 \pm 10.7$  thalamic units

(average  $\pm$  s.d. of the obtained distribution) contribute to the visually evoked thalamic current in an L4 cortical neuron. **b**, The number of unitary contributions of thalamic neurons necessary to equal the total thalamic charge as a function of the fraction of 'big contributors'. Because the units that contribute a large fraction of the total charge (big contributors) may have been under-sampled (as a consequence of a skewed distribution) we have arbitrarily increased their fraction in the pool of unitary contributions and determined the average number of unitary contributions necessary to equal the total charge, as above. Big contributors were defined as those thalamic units that contribute more than 2% of the total charge. They represent 26% of all unitary contributions in our dataset of 23 pairs (6 pairs; arrow). Increasing the fraction of big contributors (x axis) progressively reduces the average number of thalamic neurons necessary to equal the total thalamic charge evoked in response to visual stimulation (y axis). Each data point is the average  $\pm$  s.d.

**Extended Data Table 1 | Quantification of static grating-evoked thalamic excitation**

	mean $\pm$ s.d.	median	range
latency (ms)	40 $\pm$ 7	38	30 - 63
peak amplitude (pA)	105 $\pm$ 61	83	28 - 280
duration (ms)	145 $\pm$ 33	157	46 - 180

Latency (time to 20% of peak amplitude), peak amplitude and duration (time from 10–90% of integral) of thalamic excitatory postsynaptic currents (EPSCs) evoked by static gratings for each L4 neuron in Fig. 2 ( $n = 53$ ). Quantification was performed on the average response to the spatial phase eliciting the largest amplitude for each neuron.

**Extended Data Table 2 | Quantification of thalamocortical monosynaptic connections**

	mean $\pm$ s.d.	median	range
uEPSC latency (ms)	2.1 $\pm$ 0.5	2	1.4 - 3.6
uEPSC jitter ( $\mu$ s)	211 $\pm$ 47	214	107 - 289
uEPSC amplitude (pA)	10.8 $\pm$ 11.6	6.4	1.8 - 48.4
uEPSP amplitude (mV)	0.81 $\pm$ 0.79	0.57	0.14 - 3.4

uEPSC latency, jitter, amplitude and uEPSP amplitude of the 23 thalamocortical monosynaptic connections reported in the study (Methods).

# Prevalent lightning sferics at 600 megahertz near Jupiter's poles

Shannon Brown<sup>1\*</sup>, Michael Janssen<sup>1</sup>, Virgil Adumitroaie<sup>1</sup>, Sushil Atreya<sup>2</sup>, Scott Bolton<sup>3</sup>, Samuel Gulakis<sup>1</sup>, Andrew Ingersoll<sup>4</sup>, Steven Levin<sup>1</sup>, Cheng Li<sup>4</sup>, Liming Li<sup>5</sup>, Jonathan Lunine<sup>6</sup>, Sidharth Misra<sup>1</sup>, Glenn Orton<sup>1</sup>, Paul Steffes<sup>7</sup>, Fachreddin Tabataba-Vakili<sup>4</sup>, Ivana Kolmašová<sup>8,9</sup>, Masafumi Imai<sup>10</sup>, Ondřej Santolík<sup>8,9</sup>, William Kurth<sup>10</sup>, George Hospodarsky<sup>10</sup>, Donald Gurnett<sup>10</sup> & John Connerney<sup>11</sup>

**Lightning has been detected on Jupiter by all visiting spacecraft through night-side optical imaging and whistler (lightning-generated radio waves) signatures<sup>1–6</sup>. Jovian lightning is thought to be generated in the mixed-phase (liquid–ice) region of convective water clouds through a charge-separation process between condensed liquid water and water-ice particles, similar to that of terrestrial (cloud-to-cloud) lightning<sup>7–9</sup>. Unlike terrestrial lightning, which emits broadly over the radio spectrum up to gigahertz frequencies<sup>10,11</sup>, lightning on Jupiter has been detected only at kilohertz frequencies, despite a search for signals in the megahertz range<sup>12</sup>. Strong ionospheric attenuation or a lightning discharge much slower than that on Earth have been suggested as possible explanations for this discrepancy<sup>13,14</sup>. Here we report observations of Jovian lightning sferics (broadband electromagnetic impulses) at 600 megahertz from the Microwave Radiometer<sup>15</sup> onboard the Juno spacecraft. These detections imply that Jovian lightning discharges are not distinct from terrestrial lightning, as previously thought. In the first eight orbits of Juno, we detected 377 lightning sferics from pole to pole. We found lightning to be prevalent in the polar regions, absent near the equator, and most frequent in the northern hemisphere, at latitudes higher than 40 degrees north. Because the distribution of lightning is a proxy for moist convective activity, which is thought to be an important source of outward energy transport from the interior of the planet<sup>16,17</sup>, increased convection towards the poles could indicate an outward internal heat flux that is preferentially weighted towards the poles<sup>9,16,18</sup>. The distribution of moist convection is important for understanding the composition, general circulation and energy transport on Jupiter.**

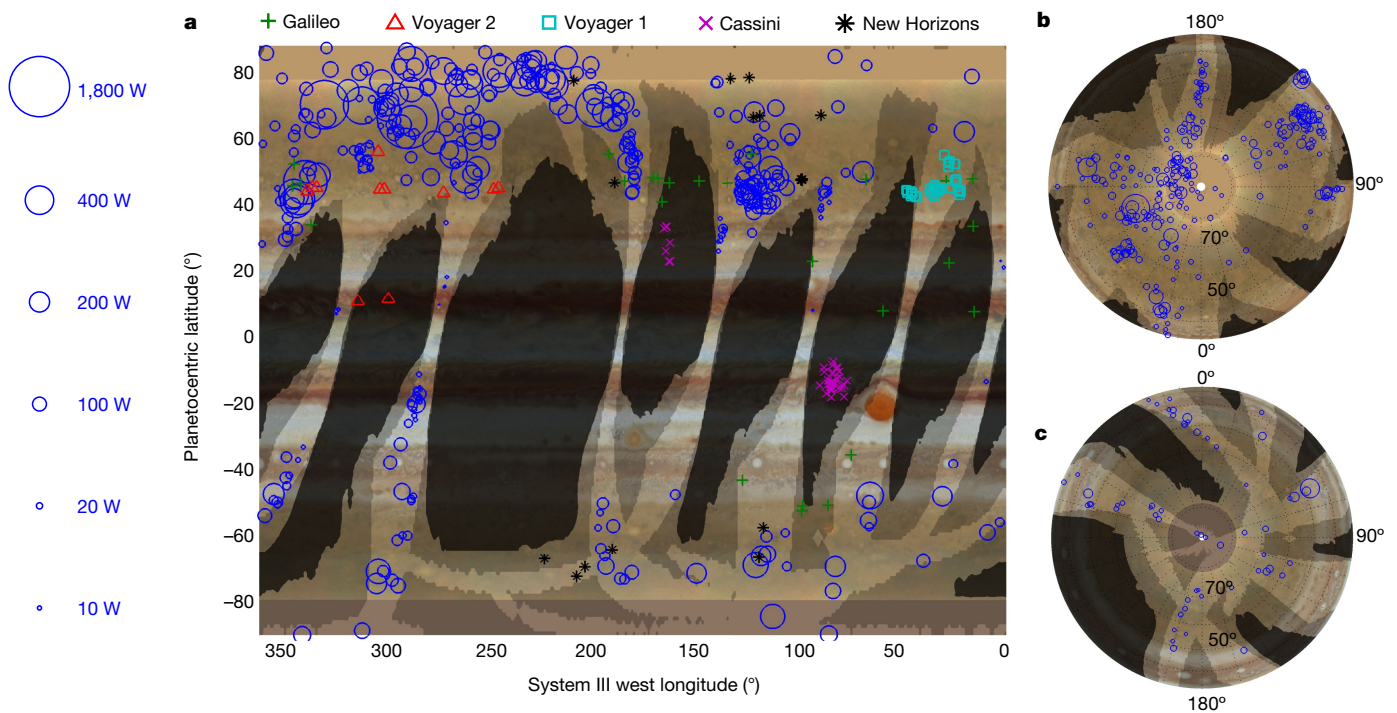
Terrestrial radio emission from lightning peaks near 1–10 kHz and falls off rapidly with the frequency  $f$ , approximately as  $f^{-4}$ , above about 10 MHz<sup>10,11</sup>. Radio emission from lightning is detectable from a spacecraft at kilohertz frequencies in the form of whistlers (lightning-generated radio waves distorted into a decreasing tone by their passage through the plasma environment of the planet), which propagate from the source to the spacecraft along magnetic field lines, and at megahertz frequencies as sferics, which propagate directly from the source to the spacecraft. On Jupiter, whistlers were previously detected by the Voyager plasma wave receiver<sup>19</sup>, but no high-frequency (10–40 MHz) sferic signals were observed by the companion planetary radio-astronomy receiver<sup>20</sup>. The Galileo probe also failed to detect high-frequency sferics<sup>1,12</sup>. One explanation for the absence of such signals is attenuation from low-altitude ionospheric layers<sup>14</sup>. However, such layers would also strongly attenuate emission at kilohertz frequencies. Therefore, a slow-discharge model with weak emission above 10 MHz has been proposed as an alternative explanation<sup>13</sup>. The closest approach of the Juno spacecraft to Jupiter is nearly 50 times greater than that of Voyager (up

to 30 dB greater signal strength), and ionospheric attenuation, which decreases as  $f^{-2}$ , is not a contributor at 600 MHz, which is the lowest-frequency channel of the Juno microwave radiometer (MWR). On the basis of modelled and measured data<sup>21</sup>, the electron density in the Jovian ionosphere is orders of magnitude lower than that required to generate even a minimally detectable ionospheric opacity at 600 MHz. Additionally, the observed variation of the MWR antenna temperature with emission angle on the planet is consistent with emission only from the deep atmosphere over all latitudes, and there is no evidence of an ionospheric contribution. The only exception is one localized spot over the portion of the aurora corresponding to the Io flux tube.

Lightning detection reveals areas of active moist convection in water clouds on Jupiter<sup>5,7,8</sup>. Our current understanding of the global distribution of lightning on Jupiter draws from limited surveys, giving an incomplete picture of the spatial distribution and frequency of moist convection. From the vantage point of Jupiter's polar orbit, the Juno observations provide new insights into the latitudinal distribution of lightning and moist convection from pole to pole. Juno is in a highly elliptical, 53-day polar orbit around Jupiter. The spacecraft is spinning at two revolutions per minute, with a spin vector roughly perpendicular to the orbit plane. The Juno MWR instrument was designed to probe thermal emission from the Jovian atmosphere well below the water-cloud region (at least 100 bar; 1 bar =  $10^5$  Pa) and thus place constraints on the deep water abundance of the planet's atmosphere. The instrument measures radiation in six microwave bands from 600 MHz to 22 GHz (1.3–50 cm)<sup>15</sup>. The two lowest-frequency channels (600 MHz and 1.26 GHz) have a Gaussian antenna pattern with a half-power width of 20°, which is scanned across the planet by the spacecraft spin.

The MWR continuously samples during Juno's orbit, integrating each radiance measurement for 0.1 s. Single positive outliers above the background atmospheric emission were observed in the time series of the 600-MHz measurement only while observing Jupiter. After eliminating all other plausible explanations for the source of these outliers, including instrument artefacts and other sources of non-thermal emission, we attribute them to lightning sferics. The lightning emission is extracted from the background signal by applying a low-pass filter to the radiometer's time series and selecting positive outliers that are six standard deviations above the noise ( $>5$  K in antenna temperature). This yields a total of 377 detections at 600 MHz through the first eight (out of the 32 planned) orbits of Juno. Each detection represents the sum of all discharges that occurred within the antenna field of view during the 0.1-s integration period. Of these, 10 MWR detections were found to be coincident in time and location with lightning whistlers detected by the Waves instrument<sup>22</sup>, further supporting lightning as the source of the

<sup>1</sup>Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA. <sup>2</sup>Climate and Space Sciences and Engineering, University of Michigan, Ann Arbor, MI, USA. <sup>3</sup>Southwest Research Institute, San Antonio, TX, USA. <sup>4</sup>California Institute of Technology, Pasadena, CA, USA. <sup>5</sup>Department of Physics, University of Houston, Houston, TX, USA. <sup>6</sup>Department of Astronomy, Cornell University, Ithaca, NY, USA. <sup>7</sup>School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, USA. <sup>8</sup>Department of Space Physics, Institute of Atmospheric Physics, The Czech Academy of Sciences, Prague, Czechia. <sup>9</sup>Faculty of Mathematics and Physics, Charles University, Prague, Czechia. <sup>10</sup>Department of Physics and Astronomy, University of Iowa, Iowa City, IA, USA. <sup>11</sup>NASA/Goddard Spaceflight Center, Greenbelt, MD, USA. \*e-mail: shannon.t.brown@jpl.nasa.gov



**Fig. 1 | MWR boresight location of each 600-MHz lightning detection during Juno's first seven successful passes.** Each MWR detection is shown as a blue circle with a diameter proportional to the minimum effective isotropic radiating power (the scale shown on the left corresponds to a). b and c show the north and south polar projection, respectively. The area of the planet surveyed by MWR during each perijove pass is

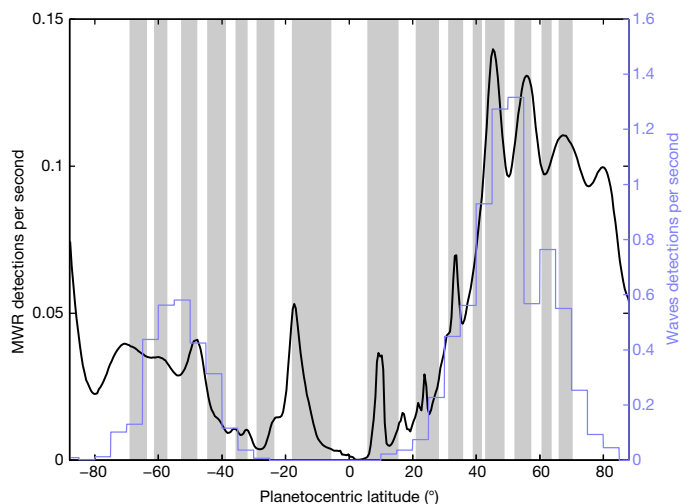
excess emission. Similar, but much fewer (12), signatures were detected at 1.2 GHz, and there were no definite detections above 1.2 GHz; this is expected, given the  $\sim f^{-4}$  dependence of lightning radio emission. The 600-MHz and 1.2-GHz antennas are on different sides of the spacecraft and therefore do not make temporally coincident observations, so a direct measure of the spectral emission slope is not possible. Additional observations are expected to provide statistical constraints, which may provide insight into the nature of the discharge process. The remainder of this paper focuses on the 600-MHz observations.

Figure 1, which illustrates the MWR boresight location and relative strength of each 600-MHz lightning detection, reveals a new and more complete picture of the global lightning distribution. The MWR detections span the locations of previous detections and show lightning polewards of 79° N, the highest latitude reported by New Horizons<sup>6</sup>. Lightning is detected at both poles but is absent near the equator. An additional notable observation is the absence of lightning in the Great Red Spot during the direct Juno overpass on 11 July 2017. The most probable source location of the lightning—the area on the planet that emitted 90% of the power received by the MWR—would fall on average within  $10^6$  km<sup>2</sup> of the boresight at the equator and  $10^9$  km<sup>2</sup> at the pole. During the first eight Juno orbits, the antenna footprint (defined by the 3-dB contour) covered  $3 \times 10^{10}$  km<sup>2</sup>, or 50% of the planet, in approximately equal amounts between the northern and southern hemispheres. Although uncertainty in the source location does not allow an exact computation of the power transmitted in the direction of the MWR, a lower bound can be computed assuming that the emission is directed at the maximum antenna gain. The derived minimum effective isotropic radiating power for lightning emission spans between 1.2 W and 1,800 W, with 87% of the detections below 200 W and 96% below 400 W. The strongest detections are preferentially weighted towards the northern hemisphere.

We derive the latitude distribution probabilistically, by spreading each detection over a latitude range weighted by the projected antenna gain and normalized by the total observation time per

indicated by the bright regions. The brightest regions show the locations of maximum gain and the less bright ones show the area covered by the 3-dB antenna pattern contour. The visible-light background image is aligned with the Great Red Spot overpass, made in July 2017. The other passes are not aligned with visible features in the image because the clouds propagate relative to the System III longitude owing to zonal winds.

latitude. Figure 2 illustrates that lightning was mostly detected in the northern mid- to high latitudes, revealing an oscillating pattern with peaks near 45° N, 56° N, 68° N and 80° N. There are also peaks centred in the North Equatorial Belt and the North Temperate Belt. The mean probability is higher in the belts (0.0045 detections per



**Fig. 2 | Lightning detections per second by the MWR and the Waves instrument as a function of latitude.** The black line shows the distribution of sferics observed by the MWR. The lightning detection locations are distributed over the latitude, which is weighted by the projected MWR antenna gain pattern. This accounts for the uncertainty in the lightning source location within the MWR beam in a probabilistic way. The blue line shows the detection frequency of whistlers by the Waves instrument as a function of magnetic footprint latitudes from the VIP4 model. The grey bars indicate the belts with the zones (white bars) in between<sup>30</sup>.

second) than in the zones (0.0035 detections per second) for absolute latitudes below 70°. Analysis of Galileo data also found more frequent lightning in belts<sup>4</sup>. In the southern hemisphere, the lightning frequency peaks at 17° S, followed by 48° S. The ambiguity of the lightning source location in the MWR beam gives an effective resolution in latitude (defined by the 3 dB antenna contour averaged over all detections) of about 2° at the equator, 5° at  $\pm 45^\circ$ , and 10° near the poles.

Considerably more lightning is detected in the northern hemisphere than in the southern hemisphere. This hemispherical asymmetry is observed in each perijove pass (spaced every 53 days), so it is not due to a single anomalously active storm in the northern hemisphere. The MWR is closer to the planet, and thus slightly more sensitive to lightning, at a given latitude in the northern hemisphere compared to the southern hemisphere. However, removing those detections in the northern hemisphere that would not have been observed in the southern hemisphere at an equivalent latitude does not affect the observed asymmetry. The equatorial zone is the only place on Jupiter with a near-zero probability of lightning detection. The boresight location of the most equatorial detection is 7.8° N. Accounting for sampling through orbit 8, equatorial lightning (within  $\pm 6^\circ$  of the equator) must occur at a rate of  $<0.03 \text{ km}^{-2} \text{ yr}^{-1}$  to have a detection probability less than 1 (or be 100 times less intense than lightning at higher latitudes). The absence of lightning at the equator is consistent with the non-detection of lightning at the Galileo probe entry site<sup>12</sup> at 6° N and the absence of visible detections within  $\pm 6^\circ$  of the equator.

Jovian rapid whistlers observed by the Waves instrument<sup>22</sup> show a similar north–south asymmetry, as indicated by the whistler detection frequency per unit time shown in Fig. 2 in 5° latitude bins. The source location of the whistlers is estimated by back-propagation to be 300 km above the 1-bar level along magnetic field lines using the VIP4 model<sup>23</sup>. The overall detection rates of whistlers are approximately ten times higher than the MWR detection rate because of the increased source power at kilohertz frequencies compared to 600 MHz<sup>10</sup>. We note also that whistlers do not show detections within about 20° of the equator, which may be explained by ducted propagation not allowing them to access the Juno altitude. High-latitude observations are mostly missing from whistler records, as these are masked by intense plasma waves in the polar-cap and auroral regions.

The moist convection distribution derived here has implications for the global water abundance and energy budget of Jupiter. While moist convection is a complex process that is influenced in part by the local water concentration, thermodynamic environment and vertical wind shear, the distribution observed here could support a preferentially poleward-weighted distribution of the outward-directed internal heat flux<sup>16,18</sup>. The high concentration of ammonia in the equatorial zone<sup>24,25</sup> suggests that the equator is close to an ideal adiabat<sup>26</sup>, potentially explaining the lack of convection at the equator. A rising air parcel in this region would have the same density as the ambient air and would not gain kinetic energy from the upward motion. Moist convection involving water clouds, as inferred from the presence of lightning, provides a constraint on the water abundance. If water were depleted globally below the value of the solar oxygen-to-hydrogen ratio, as would be concluded by taking the Galileo probe result as a global number<sup>27</sup>, there would be no liquid water<sup>28</sup> and hence lightning would be difficult to generate<sup>7,8</sup>. Moist convection models also suggest that insufficient latent heat would be available to sustain lightning-generating updrafts<sup>19,20,29</sup>. Previous studies of optical lightning imagery (and the associated moist convection) have used these arguments to suggest a global water abundance greater than the solar one, under the assumption that the convective nature of regional storms observed over a short period of time applies globally and is sustained over time<sup>5,8</sup>. The MWR lightning observations analysed here show widespread moist convective activity at nearly all latitudes consistently for over an Earth year, providing compelling evidence for a global water abundance at least as high as the solar one.

## Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0156-5>.

Received: 21 November 2017; Accepted: 14 March 2018;

Published online 6 June 2018.

- Gurnett, D. A., Shaw, R. R., Anderson, R. R., Kurth, W. S. & Scarf, F. L. Whistlers 215 observed by Voyager 1: detection of lightning on Jupiter. *Geophys. Res. Lett.* **6**, 511–514 (1979).
- Cook, A. F., Duxbury, T. C. & Hunt, G. E. First results of Jovian lightning. *Nature* **280**, 794 (1979).
- Borucki, W. J. & Magalhães, J. A. Analysis of Voyager 2 images of Jovian lightning. *Icarus* **96**, 1–14 (1992).
- Little, B. et al. Galileo images of lightning on Jupiter. *Icarus* **142**, 306–323 (1999).
- Dyudina, U. A. et al. Lightning on Jupiter observed in the H $\alpha$  line by the Cassini imaging science subsystem. *Icarus* **172**, 24–36 (2004).
- Baines, K. H. et al. Polar lightning and decadal-scale cloud variability on Jupiter. *Science* **318**, 226–229 (2007).
- Rinnert, K. Lightning on other planets. *J. Geophys. Res. D* **90**, 6225–6237 (1985).
- Gibbard, S., Levy, E. H. & Lunine, J. I. Generation of lightning in Jupiter's water cloud. *Nature* **378**, 592–595 (1995).
- Gierasch, P. J., Ingersoll, A. P., Banfield, D. & Ewald, S. P. Observation of moist convection in Jupiter's atmosphere. *Nature* **403**, 628–630 (2000).
- Oh, L. L. Measured and calculated spectral amplitude distribution of lightning sferics. *IEEE Trans. Electromagn. Compat.* **4**, 125–130 (1969).
- LeVine, D. M. & Meneghini, R. Simulation of radiation from lightning return strokes: the effects of tortuosity. *Radio Sci.* **13**, 801–809 (1978).
- Rinnert, K. et al. Measurements of radio frequency signals from lightning in Jupiter's atmosphere. *J. Geophys. Res. Planets* **103**, 22979–22992 (1998).
- Farrell, W. M. in *Radio Astronomy at Long Wavelengths* (eds Stone, R. G. et al.) 179–186 (American Geophysical Union, Washington DC, 2000).
- Zarka, P. On detection of radio bursts associated with Jovian and Saturnian lightning. *Astron. Astrophys.* **146**, L15–L18 (1985).
- Janssen, M. A. et al. MWR: microwave radiometer for the Juno mission to Jupiter. *Space Sci. Rev.* **213**, 139–185 (2017).
- Ingersoll, A. P. & Porco, C. C. Solar heating and internal heat flow on Jupiter. *Icarus* **35**, 27–43 (1978).
- Ingersoll, A. P., Gierasch, P. J., Banfield, D., Vasavada, A. R. & Galileo Imaging Team. Moist convection as an energy source for the large-scale motions in Jupiter's atmosphere. *Nature* **403**, 630–632 (2000).
- Pirraglia, J. A. Meridional energy balance of Jupiter. *Icarus* **59**, 169–176 (1984).
- Stoker, C. R. Moist convection: a mechanism for producing the vertical structure of the Jovian equatorial plumes. *Icarus* **67**, 106–125 (1986).
- Guillot, T. Condensation of methane, ammonia, and water and the inhibition of convection in giant planets. *Science* **269**, 1697–1699 (1995).
- Majeed, T., McConnell, J. C. & Gladstone, G. R. A model analysis of Galileo electron densities on Jupiter. *Geophys. Res. Lett.* **26**, 2335–2338 (1999).
- Kolmašová, I. et al. Discovery of rapid whistlers close to Jupiter implying similar lightning rates as on Earth. *Nat. Astron.* <https://doi.org/10.1038/s41550-018-0442-z> (2018).
- Connerney, J. E. P., Acuña, M. H., Ness, N. F. & Satoh, T. New models of Jupiter's magnetic field constrained by the Io flux tube footprint. *J. Geophys. Res.* **103**, 11929–11939 (1998).
- Bolton, S. J. et al. Jupiter's interior and deep atmosphere: the initial pole-to-pole passes with the Juno spacecraft. *Science* **356**, 821–825 (2017).
- Li, C. et al. The distribution of ammonia on Jupiter from a preliminary inversion of Juno Microwave Radiometer data. *Geophys. Res. Lett.* **44**, 5317–5325 (2017).
- Ingersoll, A. P. et al. Implications of the ammonia distribution on Jupiter from 1 to 100 bars as measured by the Juno microwave radiometer. *Geophys. Res. Lett.* **44**, 7676–7685 (2017).
- Niemann, H. B. et al. The composition of the Jovian atmosphere as determined by the Galileo probe mass spectrometer. *J. Geophys. Res. Planets* **103**, 22831–22845 (1998).
- Atreya, S. K. et al. Comparison of the atmospheres of Jupiter and Saturn: deep atmospheric composition, cloud structure, vertical mixing, and origin. *Planet. Space Sci.* **47**, 1243–1262 (1999).
- Hueso, R. & Sánchez-Lavega, A. A three-dimensional model of moist convection for the giant planets: the Jupiter case. *Icarus* **151**, 257–274 (2001).
- Porco, C. C. et al. Cassini imaging of Jupiter's atmosphere, satellites, and rings. *Science* **299**, 1541–1547 (2003).

**Acknowledgements** This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. The research at the University of Iowa was supported by NASA through contract 699041X with the Southwest Research Institute. The work of I.K. and O.S. was supported by grants MSM100421701 and LTAUSA17070 and by the Praemium Academiae award.

**Reviewer information** Nature thanks U. Dyudina and the other anonymous reviewer(s) for their contribution to the peer review of this work.

**Author contributions** S. Br. analysed the MWR data to find and extract the lightning observations. M.J. is the co-investigator lead of the MWR. S.A., A.I., C.L., J.L., L.L., G.O., P.S., S. Bo. and F.T.-V. contributed to the interpretation of the data and the implications for atmospheric processes. S.G., S.M. and V.A. contributed to the interpretation of the radiometric source signal. I.K., M.I. and O.S. calculated whistler rates. W.S.K., G.B.H. and D.A.G. advised on data analysis. W.S.K. is responsible for the Juno Waves instrument. J.E.P.C. provided the planetary magnetic field measurements. S. Br. wrote the manuscript with input from all authors.

**Competing interests** The authors declare no competing interests.

**Additional information**

**Extended data** is available for this paper at <https://doi.org/10.1038/s41586-018-0156-5>.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>.

**Correspondence and requests for materials** should be addressed to S.Br.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## METHODS

**Lightning detection methodology.** In the MWR data, lightning is observed as single positive outliers in the time series of the antenna temperature. The MWR integrates each sample for 0.1 s, during which time the spacecraft rotates by  $1.2^\circ$ , or about  $1/17$ th of an antenna beamwidth at 600 MHz and 1.2 GHz. This means that each sample is correlated with its neighbours, and only a short ( $<0.1$  s) radio-frequency impulse on top of the background brightness temperature distribution could produce such a discrete jump in the time series. Instrument anomalies were eliminated as the source of these outliers by evaluating high-rate data during the cruise portion of the mission and several high-rate acquisitions near apojove. No outliers greater than four standard deviations above the noise were observed in approximately 5,000 h of observations through the antenna or in the accompanying observations of internal calibration sources. During the perijove pass, outliers were observed only when the radiometer was pointed towards the planet, and not away from the planet, eliminating radiation effects or other anomalies unique to the Jovian environment as the cause. The most probable source of short radio-frequency emission bursts from Jupiter is thus lightning.

Extended Data Fig. 1 shows an example of a lightning detection. The left panel shows the antenna temperature time series from approximately one spin of the spacecraft. The centre of the plot shows the scan across Jupiter, with the peak value closest to nadir. The large lobes on either side of the planet near the limb arise from synchrotron radiation. The cold sky background forms the baseline level. The lightning observation is the single-point outlier on the planet, which is shown more clearly in the zoomed-in image in the right panel.

The outliers are extracted from the time series using a two-step process. First, the data are low-pass-filtered using a 15-point local least-squares regression smoothing method. Positive outliers in the difference between the measurements and the smoothed background that are greater than 5 K (six standard deviations above the noise floor) are identified and removed. Linear interpolation is used to fill in the gaps from the filtered outliers and a second low-pass filter is applied to this 'cleaned' time series to estimate the background antenna temperature as a function of time. Extended Data Fig. 2a shows 600-MHz antenna temperatures obtained during a single Juno spin, with the smoothed background antenna temperature overlaid. The lightning observation in this case is the red outlier point. The final step subtracts the measurements from the smoothed background antenna temperature and all positive outliers greater than 5 K are extracted as lightning observations. An example of this difference for perijove 7 is shown in Extended Data Fig. 2b. The noise on this difference is a combination of instrument noise ( $\sim 0.6$  K) and noise from the background removal. The noise (excluding the positive outliers) is consistent with a zero mean Gaussian distribution with a standard deviation of 0.8 K. The 5 K detection threshold is set conservatively to minimize the number of false positives. Extended Data Fig. 2c shows the difference between the antenna temperature and the background for all perijove passes through orbit 7. The negative-temperature part of the histogram guided the choice for the 5 K threshold, which is indicated by the vertical dashed lines in the figure.

**Analysis of detection biases.** To assess the statistical robustness of these conclusions, we must understand possible detection biases in the data. The 5 K detection threshold in the antenna temperature is based on received power, which introduces a sampling bias relative to source emission that varies with the square of the distance to the planet ( $R^2$ ). The MWR is approximately 100 times more sensitive to a single lightning flash at perijove (a few degrees north of the equator) than it is at the pole. However, the received power is the sum of all lightning flashes in a single sample, and polar observations cover 30–40 times more area on the planet than those made at the equator. Therefore, the lightning detectability in the MWR ultimately depends on how close the average transmission strength is to the threshold at a given distance and spatial density of lightning, both of which are unknown.

If we assume that lightning is sufficiently sparse so that each MWR detection originates from a single source, the difference in area from each observation is not a factor. We can therefore normalize the power received during each observation to the perijove by scaling by the square of the ratio of the observation distance to the perijove distance, as shown in Extended Data Fig. 3a. The minimum detection threshold, shown as the solid black line, varies with  $R^2$  and is slightly skewed towards the northern hemisphere because Juno's perijoves are all north of the equator. The dashed red line represents a detection threshold that is symmetric about the equator. Even if we removed the northern hemisphere observations that would not be detectable in the southern hemisphere at an equivalent

latitude, there would still be more numerous and stronger detections in the northern hemisphere. We can also compute an upper limit on the lightning rate at the equator. The MWR acquired 5,690 samples within  $\pm 6^\circ$  of the equator through perijove 8. On average, each sample is integrated over an effective area of  $2 \times 10^6 \text{ km}^2$  for 0.1 s. Therefore, for the probability of detection to be less than 1, the lightning flash rate within  $\pm 6^\circ$  must be less than  $0.03 \text{ km}^{-2} \text{ yr}^{-1}$ ,  $\frac{1}{(2 \times 10^6 \text{ km}^2)(569 \text{ s})} = 31,557,600 \text{ s yr}^{-1}$ , or be 100 times less intense than lightning at higher latitudes.

Alternatively, if we assume lightning is sufficiently dense so that the number of transmitters that are visible per observation is proportional to the area, then it is appropriate to scale the received power by the area illuminated. The signal received by the MWR from lightning is the integral of all discharges occurring in the area of the planet that is illuminated by the antenna over the 0.1 s integration period. The power at the receiver is a function of the spacecraft location and viewing angle relative to the planet. To allow comparisons between different locations on the planet, we normalize the measurements with respect to source power. However, without knowing the source location, we cannot determine the source power from the received signal. We can compute a power density to normalize the measurements. If we assume that lightning radiates isotropically with a power of  $P_{\text{iso}}$  and has a discharge frequency of  $N_l$  per unit area of the planet per unit time, then we can write the total received power as

$$P_r = \iint \frac{N_l P_{\text{iso}} G(r_{\text{dA}}) \lambda^2}{(4\pi)^2 |r_{\text{dA}}|^2} dA dt \quad (1)$$

where  $r_{\text{dA}}$  is the vector between the differential area of the planet,  $dA$ , and the MWR antenna,  $G$  is the antenna gain in that direction,  $\lambda$  is the wavelength and  $t$  is the time. We can use equation (1) to normalize the received power measurements

$$N_l P_{\text{iso}} = \frac{P_r}{\iint \frac{G(r_{\text{dA}}) \lambda^2}{(4\pi)^2 |r_{\text{dA}}|^2} dA dt} \quad (2)$$

where  $N_l P_{\text{iso}}$  is the received isotropic radiated power per unit area per unit time ( $\text{W km}^{-2} \text{ s}^{-1}$ ).

Extended Data Fig. 3b shows the lightning power normalized using equation (2). The minimum detection threshold in this case is more uniform in latitude because the illumination area increases as  $R^2$ , offsetting the decrease in signal strength with distance. In both cases, the larger number of observations in the northern hemisphere appears to be a statistically robust conclusion. Removing from the northern-hemisphere data those values that would not be detected in the southern hemisphere at an equivalent latitude gives the red line in Extended Data Fig. 4; there is negligible change in the resulting north–south asymmetry.

**Effective isotropic radiating power.** The received power is computed from the lightning antenna temperature by

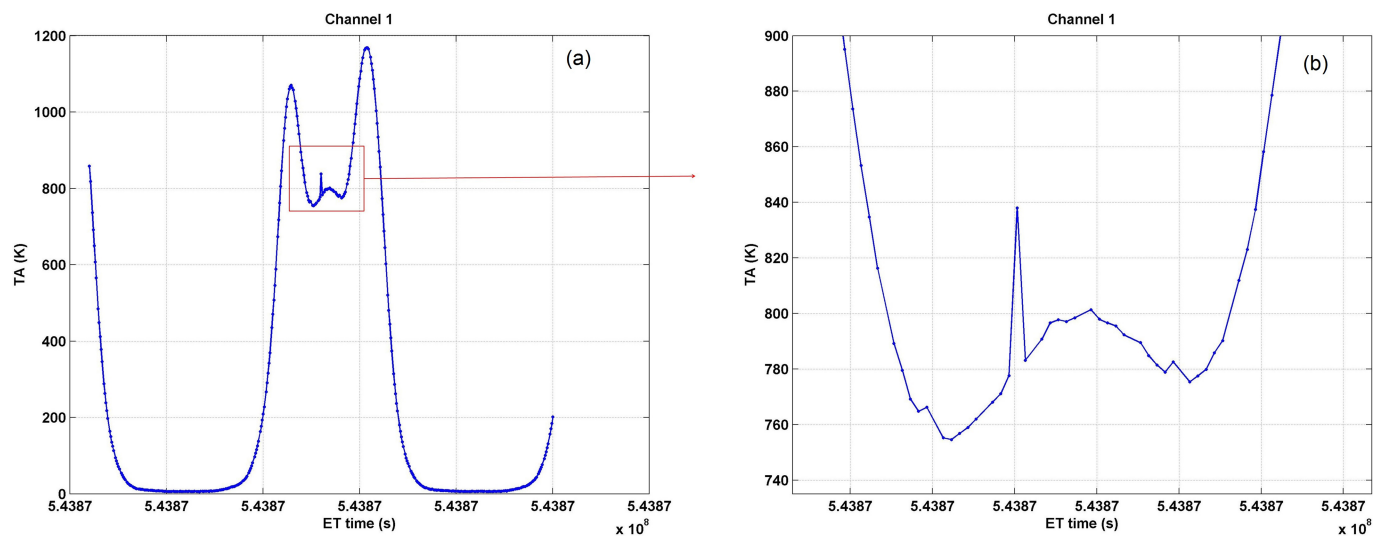
$$P_r = kBT_A \quad (3)$$

where the 600-MHz receiver bandwidth is  $B = 18 \text{ MHz}$  and  $k$  is Boltzmann's constant. Because the MWR measures a single linear polarization, if lightning emission is assumed to be unpolarized, a factor of 2 is introduced between the received and source power. The effective isotropic radiating power is then computed by

$$P_{\text{iso}} = \frac{2P_r (4\pi R)^2}{G_r \lambda^2} \quad (4)$$

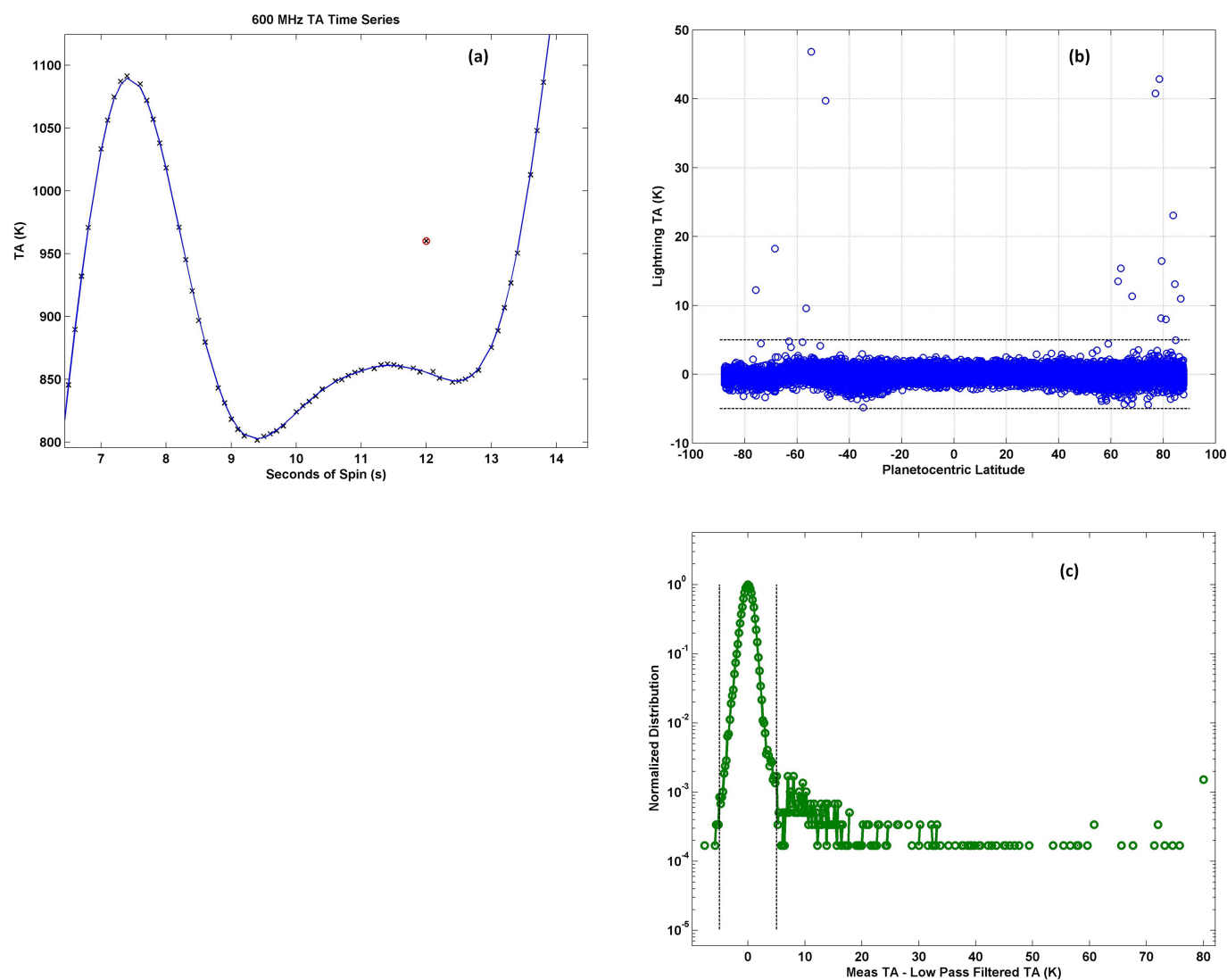
where  $R$  is taken to be the distance of the boresight vector to Jupiter,  $G_r$  is the maximum antenna gain (19.77 dB) and the wavelength  $\lambda$  is 0.5 m. The result represents the minimum radiated power and only applies when the boresight is pointed directly at the lightning. The lightning is probably detected over a broad range of angles (and hence antenna gains) relative to the boresight.

**Data availability.** The Juno MWR data that support the findings of this study are available from the Planetary Data System archive (<https://pds.nasa.gov/index.shtml>) as 'Juno Jupiter MWR reduced data records v1.0' (dataset JNO-J-MWR-3-RDR-V1.0).



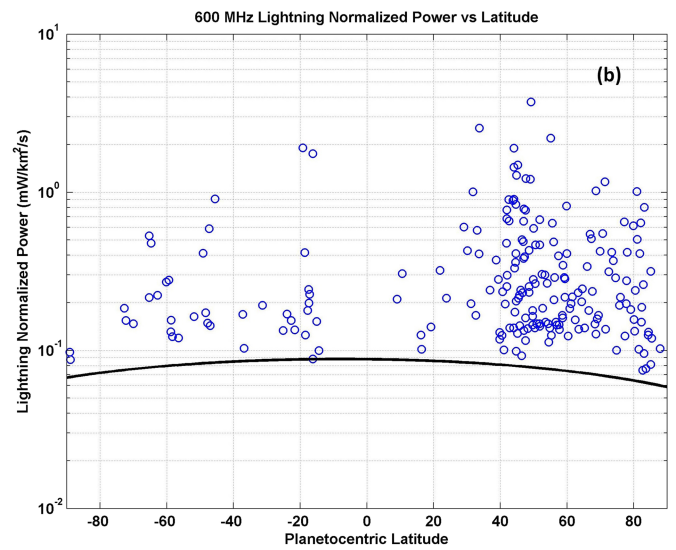
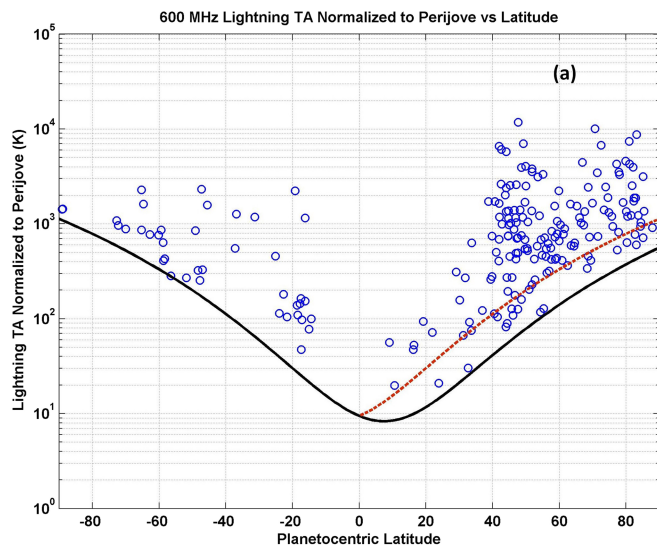
**Extended Data Fig. 1 | Example of lightning detection in the MWR antenna temperature time series. a,** Antenna temperature measurements obtained during a spin of the spacecraft. The scan from limb to limb of

Jupiter is shown at the centre of the image and enlarged in panel **b**. The single positive outlier is the additive emission from the lightning discharge above the background emission from the atmosphere.



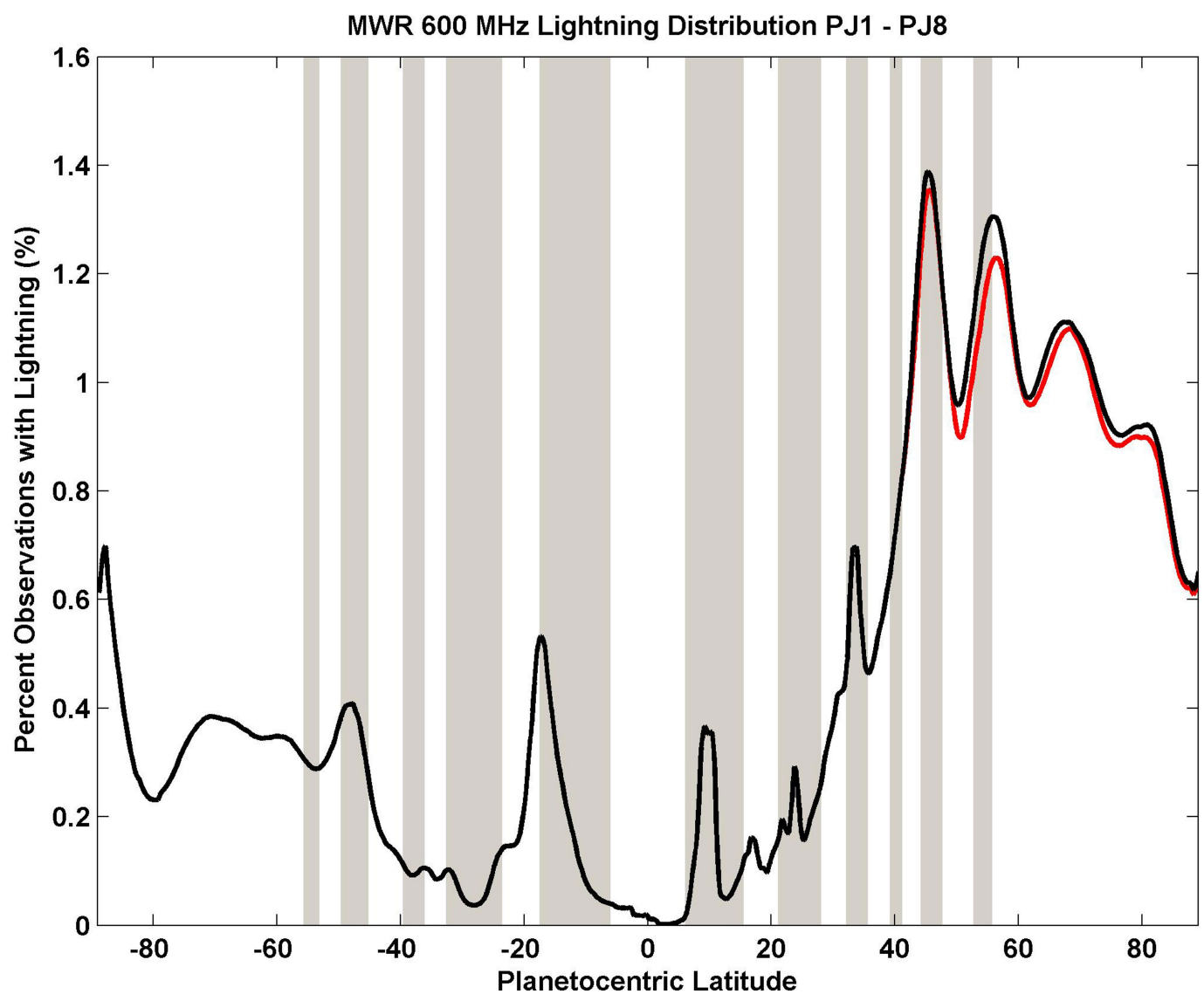
**Extended Data Fig. 2 | Illustration of the lightning extraction process.**  
**a,** Smoothed background antenna temperature (TA) and the data.  
**b,** Difference between the measurements and the smoothed background antenna temperature for perijove 7. **c,** Differences between the MWR data

and the background antenna temperature for all perijoves. The dotted lines in **b** and **c** indicate where the detection threshold is set relative to the variance in the data.



**Extended Data Fig. 3 | Normalized 600-MHz lightning power, expressed as antenna temperature, as a function of latitude. a,** Power normalized to the perijove distance by the square of the distance. This normalization is used if the observed power from each detection originates from a single source, which is expected for discharge rates less than  $300 \text{ km}^{-2} \text{ yr}^{-1}$  near

the equator and  $0.3 \text{ km}^{-2} \text{ yr}^{-1}$  at the poles. **b,** Power normalized by both the distance and the area covered by the antenna pattern, which is used if the observed power originates from several sources and should be scaled per unit area.



**Extended Data Fig. 4 | Lightning detections per second by the MWR and the Waves instrument as a function of latitude. The same MWR**

distribution as that shown in Fig. 2, but with the red line showing the distribution with an equalized detection threshold as a function of latitude.

# A per-cent-level determination of the nucleon axial coupling from quantum chromodynamics

C. C. Chang<sup>1,2</sup>, A. N. Nicholson<sup>1,3,4</sup>, E. Rinaldi<sup>1,5,6</sup>, E. Berkowitz<sup>6,7</sup>, N. Garron<sup>8</sup>, D. A. Brantley<sup>1,6,9</sup>, H. Monge-Camacho<sup>1,9</sup>, C. J. Monahan<sup>10,11</sup>, C. Bouchard<sup>9,12</sup>, M. A. Clark<sup>13</sup>, B. Joó<sup>14</sup>, T. Kurth<sup>1,15</sup>, K. Orginos<sup>9,16</sup>, P. Vranas<sup>1,6</sup> & A. Walker-Loud<sup>1,6\*</sup>

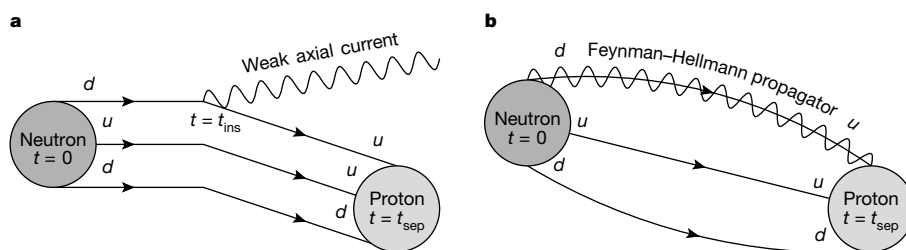
The axial coupling of the nucleon,  $g_A$ , is the strength of its coupling to the weak axial current of the standard model of particle physics, in much the same way as the electric charge is the strength of the coupling to the electromagnetic current. This axial coupling dictates the rate at which neutrons decay to protons, the strength of the attractive long-range force between nucleons and other features of nuclear physics. Precision tests of the standard model in nuclear environments require a quantitative understanding of nuclear physics that is rooted in quantum chromodynamics, a pillar of the standard model. The importance of  $g_A$  makes it a benchmark quantity to determine theoretically—a difficult task because quantum chromodynamics is non-perturbative, precluding known analytical methods. Lattice quantum chromodynamics provides a rigorous, non-perturbative definition of quantum chromodynamics that can be implemented numerically. It has been estimated that a precision of two per cent would be possible by 2020 if two challenges are overcome<sup>1,2</sup>: contamination of  $g_A$  from excited states must be controlled in the calculations and statistical precision must be improved markedly<sup>2–10</sup>. Here we use an unconventional method<sup>11</sup> inspired by the Feynman–Hellmann theorem that overcomes these challenges. We calculate a  $g_A$  value of  $1.271 \pm 0.013$ , which has a precision of about one per cent.

To demonstrate the efficacy of lattice quantum chromodynamics (LQCD) for nuclear physics research, one must begin by demonstrating control over the simplest quantities, such as  $g_A$ . In addition to those mentioned above, there are a number of challenges in using LQCD to compute properties of nucleons and nuclei. The first challenge arises from the non-perturbative features of quantum chromodynamics

(QCD) itself. QCD describes the interactions between quarks and gluons, the basic constituents of nucleons, through the Lagrangian density  $L_{\text{QCD}} = -G^2/(4g) + \sum_q \bar{\psi}_q (D + m_q) \psi_q$ , where the quark fields,  $\psi_q$ , come in flavours  $q = \{u, d, s, \dots\}$  with masses  $m_q = \{m_u, m_d, m_s, \dots\}$ .  $G^2$  describes the nonlinear gluon self-interactions and  $D$  includes the quark–gluon interactions, both with a strength determined by the coupling,  $g$ . Most of nuclear physics depends on only three or four input parameters from QCD:  $g$ , the light-quark masses,  $m_u$  and  $m_d$ , and in some cases the strange-quark mass,  $m_s$ . Once these parameters are fixed, and electroweak corrections are added, all of nuclear physics—from the kiloelectronvolt energy levels in nuclei to the energy densities of the neutron star equation of state (a few hundred megaelectronvolts per cubic fermi (fm), where  $1 \text{ fm} = 10^{-15} \text{ m}$ )—can in principle be predicted from QCD.

At short distances (high energies), such as those explored by the Large Hadron Collider at CERN, QCD has been rigorously tested, because in this energy regime  $g \ll 1$  and perturbative methods are applicable. At long distances of approximately 1 fm (low energies), which are characteristic of nuclear physics,  $g$  is large and perturbation theory fails to converge. Consequently, quarks and gluons are confined in protons, neutrons and other hadrons observed experimentally. Fortunately, non-perturbative calculations can be carried out in the strong-coupling regime using LQCD, the only first-principles approach known to control all sources of systematic uncertainty.

LQCD is the formulation of QCD on a finite four-dimensional space-time lattice, following the Feynman path-integral description. Monte Carlo methods are used to sample the resulting high-dimensional integrals stochastically. The values of the lattice spacing,  $a$ , and finite size,



**Fig. 1 | Feynman diagrams of  $g_A$ .** The decay of a neutron to a proton occurs when one of the down quarks ( $d$ ) in the neutron is converted to an up quark ( $u$ ) via the vector and axial components of the weak current. Not depicted in these figures are the infinite set of diagrams describing the coupling of gluons to the quarks and of gluons to gluons and the dynamical production and annihilation of quark–anti-quark pairs. Because of this infinite set of graphs, the use of a computational approach

to QCD is required. The time,  $t$ , refers to calculational details discussed in the text. **a**, The standard method of computing  $g_A$  relies on three different times, the creation time,  $t = 0$ , the current insertion time,  $t_{\text{ins}}$ , and the separation time,  $t_{\text{sep}}$ . Controlling the excited state systematics requires varying both  $t_{\text{ins}}$  and  $t_{\text{sep}}$ . **b**, Our Feynman–Hellmann method<sup>11</sup> sums over all possible interaction times ( $t_{\text{ins}}$ ) of the external weak axial current, leading to an exponential enhancement of the signal.

<sup>1</sup>Nuclear Science Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. <sup>2</sup>Interdisciplinary Theoretical and Mathematical Sciences (iTHEMS) Program, RIKEN, Saitama, Japan.

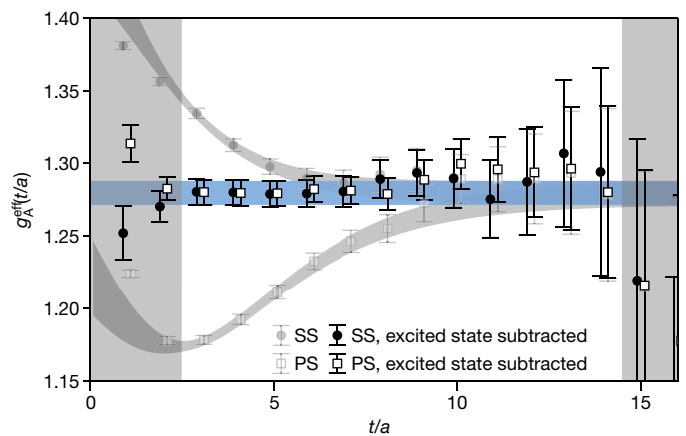
<sup>3</sup>Department of Physics, University of California, Berkeley, CA, USA. <sup>4</sup>Department of Physics and Astronomy, University of North Carolina, Chapel Hill, NC, USA. <sup>5</sup>RIKEN-BNL Research Center, Brookhaven National Laboratory, Upton, NY, USA. <sup>6</sup>Physics Division, Lawrence Livermore National Laboratory, Livermore, CA, USA. <sup>7</sup>Institut für Kernphysik und Institute for Advanced Simulation, Forschungszentrum Jülich, Jülich, Germany. <sup>8</sup>Theoretical Physics Division, Department of Mathematical Sciences, University of Liverpool, Liverpool, UK. <sup>9</sup>Department of Physics, The College of William and Mary, Williamsburg, VA, USA. <sup>10</sup>Physics Department and Astronomy, Rutgers, The State University of New Jersey, Piscataway, NJ, USA. <sup>11</sup>Institute for Nuclear Theory, University of Washington, Seattle, WA, USA. <sup>12</sup>School of Physics and Astronomy, University of Glasgow, Glasgow, UK. <sup>13</sup>NVIDIA Corporation, Santa Clara, CA, USA. <sup>14</sup>Scientific Computing Group, Thomas Jefferson National Accelerator Facility, Newport News, VA, USA. <sup>15</sup>NERSC, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. <sup>16</sup>Theory Center, Thomas Jefferson National Accelerator Facility, Newport News, VA, USA. \*e-mail: [awalker-loud@lbl.gov](mailto:awalker-loud@lbl.gov)

$L$ , are chosen to encompass the characteristic length scales emergent from QCD, such as the proton radius,  $r_p \approx 0.8$  fm. In present calculations, typical lattice spacings are  $0.04 \lesssim a/r_p \lesssim 0.2$  and typical spatial extents are  $3 \lesssim L/r_p \lesssim 8$ . The continuum and infinite-volume limits are recovered through an extrapolation with several values of  $a$  and  $L$ . Additionally, the input values of the quark masses in LQCD calculations do not typically reproduce their physical values. State-of-the-art calculations now regularly obtain  $m_\pi^{\text{calc.}} \approx 130$  MeV, where the pion ( $\pi$ ), the lightest hadron, is used to calibrate the input light-quark mass according to the experimental value<sup>12</sup>  $m_\pi^{\text{exp.}} = 139.6$  MeV. In this work,  $m_\pi$  ranges between 130 MeV and 400 MeV, allowing for a fully controlled interpolation (the input parameters of the calculation are provided in Extended Data Table 2). The continuum and infinite-volume extrapolations and pion-mass interpolations are necessary for all LQCD calculations to compare the results with experimental values and make predictions.

Over the past decade, the LQCD community has determined hadronic properties for mesons (which are QCD eigenstates composed of one quark and one antiquark) with fully controlled systematic uncertainties at the sub-per-cent level, yielding some of the most stringent tests of the standard model. The Flavour Lattice Averaging Group produces a world average of meson properties determined from LQCD<sup>13</sup>, similar to the Particle Data Group's (PDG) averages of experimental results<sup>12</sup>. In contrast to mesons, stochastic sampling of the nucleon path-integral results in an exponentially smaller signal-to-noise ratio, hence requiring exponentially more computational resources to replicate the precision achieved for meson properties. In fact, only a handful of LQCD calculations involving nucleons have demonstrated control over all sources of uncertainty<sup>14,15</sup>. Insight provided by previous LQCD calculations of  $g_A$  also identifies the contamination of excited states as another major source of uncertainty<sup>2–10</sup>. Excited state contamination, which results from the imperfect coupling of the chosen creation operators to the state of physical interest, is present in all lattice calculations and has been proven to be particularly problematic for calculations of  $g_A$ .

The standard method of calculating  $g_A$ , as shown in Fig. 1a, relies on two independent separation times: the separation time of the neutron and proton ( $t_{\text{sep}}$ ) and the separation time of the neutron and the insertion of the weak axial current ( $t_{\text{ins}}$ ). Although  $g_A$  is independent of both times, the contamination from excited states is time-dependent. This contamination shifts the calculated values of  $g_A$  at small time separations, but vanishes exponentially with respect to  $t_{\text{ins}}$  and  $t_{\text{sep}} - t_{\text{ins}}$ . Computational limitations restrict the calculations to fixed (and relatively small) neutron–proton separation times, requiring multiple calculations with varying values of  $t_{\text{sep}}$  to fully control the excited-state contributions. However, the relative stochastic noise grows exponentially with  $t_{\text{sep}}$ , while it only vanishes with the square root of the stochastic sample size. Therefore, overcoming this noise requires exponentially more computational resources, rendering the standard method an expensive strategy.

By contrast, the method that we use in this work<sup>11</sup>, which is inspired by the Feynman–Hellmann theorem, uses an explicit sum over all current insertion times,  $t_{\text{ins}}$  (Fig. 1b), with the ability to vary  $t = t_{\text{sep}}$ , at the numerical cost of a single separation time of the standard method: all excited state contributions depend only on  $t$  and the computation must approach  $g_A$  asymptotically in the large- $t$  limit (Fig. 2). By analysing the spectrum and  $g_A$  matrix element calculations simultaneously with nonlinear regression, we demonstrate the ability to fully control excited state contributions and determine precise values of  $g_A$ , as suggested by the agreement between the data (grey points with error bars) and the fit ansatzes (grey bands). In Supplementary Information (section S.4 and Figs. 9–15) and in Extended Data Fig. 1, we show that this is true for all ensembles (different choices of  $a$ ,  $L$  and  $m_\pi$ ) used in our calculation. In summary, this Feynman–Hellman-theorem-inspired method<sup>11</sup> provides access to more data ( $t = t_{\text{sep}}$  in Fig. 2) with a reduced computational cost, allowing us to remove the unwanted excited state contamination and utilize data at early separation times, where the signal-to-noise ratio is exponentially more precise, thus resolving both of the aforementioned major challenges in determining  $g_A$ .



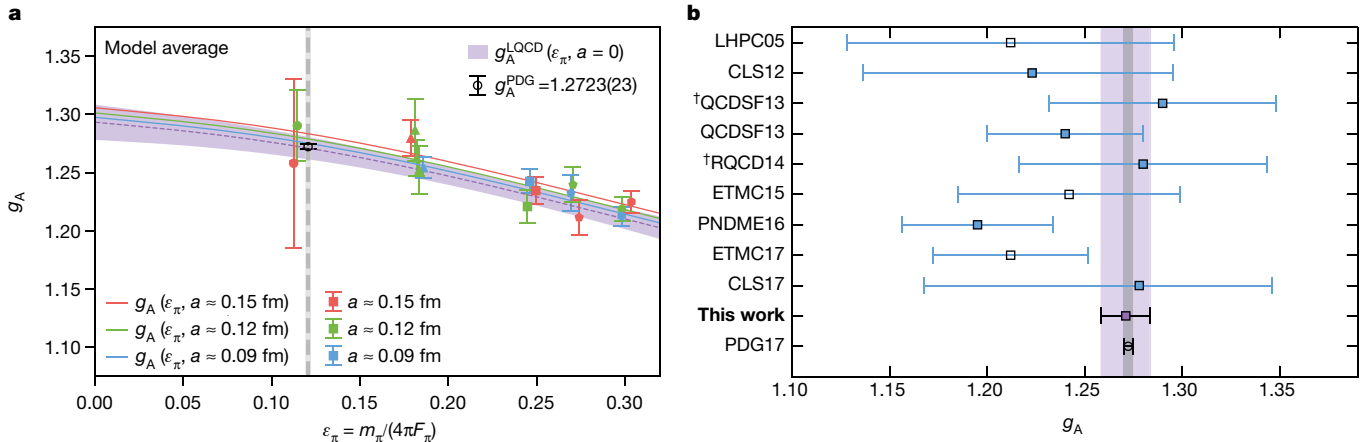
**Fig. 2 | Demonstration of the improved method<sup>11</sup> on an ensemble with lattice spacing  $a \approx 0.09$  fm and  $m_\pi \approx 220$  MeV.** The two sets of results for  $g_A^{\text{eff}}(t/a)$  correspond to different choices of annihilation operators for the nucleon, denoted as ‘SS’ and ‘PS’. At long times, both values must approach the ground state value of  $g_A$  asymptotically, whereas at short times, they couple differently to the excited state contributions. The raw numerical results are shown in grey and the grey bands represent the full fit to the data (points inside the vertical grey bands are not included in the fits). Error bars correspond to one standard error of the mean (s.e.m.). The solid black and white data points are reconstructed from the two datasets, with the excited states (determined by the fit) subtracted from the raw results. The solid blue band is the ground state value of  $g_A$ , determined by the full fit. We make efficient use of points at small Euclidean times, before the stochastic noise overwhelms the signal. The agreement between the subtracted data and the asymptotic large-time value of  $g_A$ , even at short times, demonstrates our control over excited state contributions. The time axis is given in dimensionless lattice units, with  $a \approx 0.09$  fm corresponding to  $3 \times 10^{-25}$  s, so that  $t/a = 2$  corresponds to  $6 \times 10^{-25}$  s.

What remains is to extrapolate the values of  $g_A$  obtained from our lattice calculations to the physical parameters. Effective field theory<sup>16</sup> (EFT) is employed to provide a rigorous prescription for performing the continuum and infinite-volume extrapolations, along with the interpolation to the physical pion mass. First, one identifies the relevant degrees of freedom for low-energy nuclear physics, which are the nucleons and pions. Second, one identifies a small expansion parameter,  $\varepsilon$ , which often emerges through a ratio of length scales; for pions, this is  $\varepsilon_\pi = m_\pi/(4\pi F_\pi)$ , where  $F_\pi$  is a quantity known as the pion decay constant.  $F_\pi^{\text{exp.}}$  has been measured<sup>12</sup> to be 92.1(1.2) MeV and  $\varepsilon_\pi^{\text{exp.}} \approx 0.12$ . The resulting effective field theory may be systematically improved: when working to  $O(\varepsilon^n)$  (where  $n = 0$  denotes leading order,  $n = 1$  next-to-leading order, and so on), the truncation errors enter at  $O(\varepsilon^{n+1})$ .

Chiral perturbation theory is the effective field theory of pions<sup>17</sup> and their interactions with nucleons<sup>18</sup>, and it describes all possible interactions between them that are consistent with the symmetries of QCD, ordered by increasing powers of  $\varepsilon_\pi$ . Although the forms of the interactions are known, the strengths of the interactions are emergent low-energy couplings and can be determined only from experiments or LQCD calculations. However, once the couplings are known, chiral perturbation theory can be used to calculate new quantities and can be used to describe the simulated universes where the quark masses differ from those in nature. This allows for a model-independent interpolation of LQCD results to  $m_\pi^{\text{exp.}}$ .

Chiral perturbation theory is also extended to account for artefacts arising from the finite volume of the lattice<sup>19</sup>. For the large volumes used in our calculation, the small parameter controlling the finite-volume corrections scales approximately as  $\varepsilon_L = e^{-m_\pi L}$ . Extended Data Fig. 2 shows consistency between the predicted finite-volume corrections and our results at fixed pion mass.

Artefacts introduced by our calculation at non-zero lattice spacing are also accounted for using effective field theory. Unlike the



**Fig. 3 | Physical-point extrapolation for this work and summary of  $g_A$  values calculated with LQCD. a**, The solid red, green and blue curves are the central values of  $g_A$  as a function of  $\epsilon_\pi$  at fixed lattice spacings and infinite volume, and the black circle represents the experimental value. The magenta band represents the central 68%-confidence band of the continuum and infinite-volume extrapolated value of  $g_A$  as a function of  $\epsilon_\pi$ , and its range at the physical pion mass, given by its intersection with the grey band, is our main result. The numerical results have been adjusted to their infinite-volume values. Some of the results are slightly shifted horizontally for visual clarity. **b**, Summary of selected LQCD calculations of  $g_A$ , the result of this work and the experimental result

dependence on  $\epsilon_\pi$  and  $\epsilon_L$ , which is governed purely by the long-distance dynamics of QCD, the continuum extrapolation depends on the specific discretization of the QCD Lagrangian, or lattice action, employed in the calculation. To parameterize these artefacts, one uses Symanzik's effective field theory<sup>20</sup> and expands the non-local discretized action around small lattice spacings, which gives a series of purely local interactions, allowing their effects in low-energy dynamics to be systematically incorporated. The dependence on the choice of discretization must vanish in the continuum limit because the only interactions remaining are those of QCD. The lattice action that we have chosen<sup>21</sup> was designed to minimize the leading discretization errors so that the leading corrections scale as  $O(a^2)$  and also to preserve more of the underlying symmetries of QCD. This choice of lattice action yields a mild continuum extrapolation (Extended Data Fig. 3).

The final extrapolation of our results (Extended Data Table 1) is presented in Fig. 3a. For quantities with mild pion-mass dependence, such as  $g_A$ , a simple Taylor expansion in terms of  $\epsilon_\pi$  or  $\epsilon_\pi^2$ , in addition to the chiral perturbation theory extrapolation, provides a robust extrapolation of the results. We perform the extrapolation with several models and our final result is determined as a model average, depicted in Extended Data Fig. 4 and described in detail in Supplementary Information (sections S.6 and S.7A). Our final result,  $g_A = 1.271 \pm 0.013$ , with the uncertainties broken down to the different contributions of statistical (s), chiral ( $\chi$ ), continuum (a), infinite volume (v), isospin breaking (i) and model selection (m) is

$$g_A = 1.2711(103)^s(39)^{\chi}(15)^a(19)^v(04)^i(55)^m \quad (1)$$

This value is commensurate with the experimentally determined value<sup>12,22–28</sup>,  $g_A^{\text{PDG}} = 1.2723(23)$ .

Figure 3b summarizes the improvement in the LQCD determination of  $g_A$  achieved by this work. These results are derived using three lattice spacings, five values of the pion (quark) mass and multiple volumes, which control the three standard extrapolations (the input values of the parameters used in our calculation are provided in Extended Data Table 2). Additionally, we demonstrate that our result is robust under different truncations and variations in the extrapolation function (Extended Data Fig. 5) and that the perturbative expansion converges

(PDG17). The vertical magenta band shows our full uncertainty, and the vertical grey band is the experimental uncertainty. Results with closed symbols include an extrapolation to the continuum limit, whereas results with open symbols include only an extrapolation or interpolation to the physical pion mass. When provided separately, the statistical and systematic uncertainties are added in quadrature. Labels marked with <sup>†</sup> were indirectly obtained from an extrapolation of  $g_A/F_\pi$ . The displayed results are from LHPC05<sup>3</sup>, CLS12<sup>4</sup>, QCDSF13<sup>5</sup>, RQCD14<sup>6</sup>, ETMC15<sup>7</sup>, PNDME16<sup>2</sup>, ETMC17<sup>8</sup> and CLS17<sup>9</sup>. The averaged experimental determination is obtained from the PDG12. Uncertainties are one s.e.m.

over the range of parameters used, as discussed in Supplementary Information (section S.7A) and shown in Extended Data Fig. 6. Details about individual contributions to our total uncertainty can be found in Supplementary Information, section S.7B.

Our result, equation (1), is predominantly limited by statistics. This signifies a straightforward path to further improvement: more precise results at the physical pion mass will reduce the statistical, chiral extrapolation and model-selection uncertainties, which are the three largest. An uncertainty comparable to that of measurements may offer insight into the upward-trending value of  $g_A$  observed in the most recent set of experiments<sup>12</sup>. At present, our result has a noticeable phenomenological impact, as depicted in Extended Data Fig. 7. Using effective field theory, experimental results from collider and low-energy experiments can be used to place bounds on right-handed beyond-standard-model currents<sup>29</sup>, with our result placing one of the most stringent bounds.

### Code and data availability

The majority of the software used for this work is publicly available on <https://www.github.com>. The Chroma software suite (developed by USQCD) is available at <https://github.com/JeffersonLab/chroma>. This is linked with QUDA (the optimized library for performing calculations on NVIDIA GPU-enabled machines), which is available at <https://github.com/lattice/quda>. A private C++ layer is compiled on top of Chroma, which will be available at <https://github.com/callat-qcd/lalibe>. The original LQCD correlation functions, as well as our analysis results, are available at [https://github.com/callat-qcd/project\\_gA](https://github.com/callat-qcd/project_gA) and <https://zenodo.org/record/1241374>.

Received: 24 July 2017; Accepted: 28 February 2018;

Published online: 30 May 2018

- Detmold, W. *Cold Nuclear Physics, USQCD Annual Progress Review to the US DOE* [https://www.usqcd.org/reviews/June2016Review/Talks/LQCD%20Rev16-USQCD\\_ColdNuclear\\_2016\\_Detmold.pdf](https://www.usqcd.org/reviews/June2016Review/Talks/LQCD%20Rev16-USQCD_ColdNuclear_2016_Detmold.pdf) (2016).
- Bhattacharya, T. et al. Axial, scalar and tensor charges of the nucleon from 2+1+1-flavor lattice QCD. *Phys. Rev. D* **94**, 054508 (2016).
- LHPC Collaboration. Nucleon axial charge in full lattice QCD. *Phys. Rev. Lett.* **96**, 052001 (2006).
- Capitani, S. et al. The nucleon axial charge from lattice QCD with controlled errors. *Phys. Rev. D* **86**, 074502 (2012).
- Horsley, R. et al. Nucleon axial charge and pion decay constant from two-flavor lattice QCD. *Phys. Lett. B* **732**, 41–48 (2014).

6. Bali, G. S. et al. Nucleon isovector couplings from  $N_f = 2$  lattice QCD. *Phys. Rev. D* **91**, 054501 (2015).
7. Abdel-Rehim, A. et al. Nucleon and pion structure with lattice QCD simulations at physical value of the pion mass. *Phys. Rev. D* **92**, 114513 (2015); erratum **93**, 039904 (2016).
8. Alexandrou, C. et al. Nucleon axial form factors using  $N_f = 2$  twisted mass fermions with a physical value of the pion mass. *Phys. Rev. D* **96**, 054507 (2017).
9. Capitani, S. et al. Iso-vector axial form factors of the nucleon in two-flavour lattice QCD. Preprint at <https://arxiv.org/abs/1705.06186> (2017).
10. Lin, H.-W. et al. Parton distributions and lattice QCD calculations: a community white paper. *Prog. Part. Nucl. Phys.* **100**, 107–160 (2018).
11. Bouchard, C., Chang, C. C., Kurth, T., Orginos, K. & Walker-Loud, A. On the Feynman–Hellmann theorem in quantum field theory and the calculation of matrix elements. *Phys. Rev. D* **96**, 014504 (2017).
12. Particle Data Group. Review of particle physics. *Chin. Phys. C* **40**, 100001 (2016).
13. Aoki, S. et al. Review of lattice results concerning low-energy particle physics. *Eur. Phys. J. C* **74**, 2890 (2017).
14. Durr, S. et al. Ab initio determination of light hadron masses. *Science* **322**, 1224–1227 (2008).
15. Borsanyi, S. et al. Ab initio calculation of the neutron-proton mass difference. *Science* **347**, 1452–1455 (2015).
16. Weinberg, S. Phenomenological Lagrangians. *Physica A* **96**, 327–340 (1979).
17. Gasser, J. & Leutwyler, H. Chiral perturbation theory to one loop. *Ann. Phys.* **158**, 142–210 (1984).
18. Jenkins, E. E. & Manohar, A. V. Baryon chiral perturbation theory using a heavy fermion Lagrangian. *Phys. Lett. B* **255**, 558–562 (1991).
19. Gasser, J. & Leutwyler, H. Spontaneously broken symmetries: effective Lagrangians at finite volume. *Nucl. Phys. B* **307**, 763–778 (1988).
20. Symanzik, K. Continuum limit and improved action in lattice theories: (I). Principles and  $\phi^4$  theory. *Nucl. Phys. B* **226**, 187–204 (1983).
21. Berkowitz, E. et al. Möbius domain-wall fermions on gradient-flowed dynamical HISQ ensembles. *Phys. Rev. D* **96**, 054513 (2017).
22. Bopp, P. et al. Beta-decay asymmetry of the neutron and  $g_A/g_V$ . *Phys. Rev. Lett.* **56**, 919–922 (1986).
23. Yerozolimsky, B., Kuznetsov, I., Mostovoy, Y. & Stepanenko, I. Corrigendum: corrected value of the beta-emission asymmetry in the decay of polarized neutrons measured in 1990. *Phys. Lett. B* **412**, 240–241 (1997).
24. Liaud, P. et al. The measurement of the beta asymmetry in the decay of polarized neutrons. *Nucl. Phys. A* **612**, 53–81 (1997).
25. Mostovoi, Yu. A. et al. Experimental value of  $G_A/G_V$  from a measurement of both P-odd correlations in free-neutron decay. *Phys. At. Nucl.* **64**, 1955–1960 (2001) [transl.]; *Yad. Fiz.* **64**, 2040–2045 (2001).
26. Schumann, M. et al. Measurement of the proton asymmetry parameter in neutron beta decay. *Phys. Rev. Lett.* **100**, 151801 (2008).
27. Mund, D. et al. Determination of the weak axial vector coupling  $\lambda = g_A/g_V$  from a measurement of the  $\beta$ -asymmetry parameter  $A$  in neutron beta decay. *Phys. Rev. Lett.* **110**, 172502 (2013).
28. UCNA Collaboration. Precision measurement of the neutron  $\beta$ -decay asymmetry. *Phys. Rev. C* **87**, 032501 (2013).
29. Alioli, S., Cirigliano, V., Dekens, W., de Vries, J. & Mereghetti, E. Right-handed charged currents in the era of the Large Hadron Collider. *J. High Energy Phys.* **5**, 86 (2017).
30. Bazavov, A. et al. (MILC). Lattice QCD ensembles with four flavors of highly improved staggered quarks. *Phys. Rev. D* **87**, 054505 (2013).

**Acknowledgements** We thank C. Bernard, A. Bernstein, P. J. Bickel, C. Detar, A. X. El-Khadra, W. Haxton, Y. Hsia, V. Koch, A. S. Kronfeld, W. T. Lee, G. P. Lepage, E. Mereghetti, G. Miller, A. E. Raftery, D. Toussaint and F. Yuan for discussions. We thank E. Mereghetti for the updated Extended Data Fig. 7<sup>29</sup>. We thank the MILC Collaboration for providing their highly improved staggered quark configurations<sup>30</sup> without restriction. Computer time was awarded to CalLat (2016) by the Innovative and Novel Computational Impact on Theory and Experiment (INCITE) programme, as well as by the Lawrence Livermore National Laboratory (LLNL) Multiprogrammatic and Institutional Computing programme through a Tier-1 Grand Challenge award. This research used the NVIDIA GPU-accelerated Titan supercomputer at the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Office of Science of the US Department of Energy under contract number DE-AC05-00OR22725, the GPU-enabled Surface and RZHasGPU clusters, and Vulcan, a BG/Q supercomputer, all at LLNL. This work was supported by the NVIDIA Corporation (M.A.C.), the DFG and the NSFC Sino-German CRC110 (E.B.), an LBNL LDRD (A.W.-L.), the RIKEN Special Postdoctoral Researcher Program (E.R.), the Leverhulme Trust (N.G.), the US Department of Energy, Office of Science: Office of Nuclear Physics (E.B., C.B., D.A.B., C.C.C., T.K., C.J.M., H.M.-C., A.N.N., E.R., B.J., K.O., P.V. and A.W.-L.); Office of Advanced Scientific Computing (E.B., B.J., T.K. and A.W.-L.); Nuclear Physics Double Beta Decay Topical Collaboration (D.A.B., H.M.-C. and A.W.-L.); and the DOE Early Career Award Program (D.A.B., C.C.C., H.M.-C. and A.W.-L.). This work (E.B., E.R. and P.V.) was performed under the auspices of the US Department of Energy by LLNL under contract number DE-AC52-07NA27344. Part of this work was performed at the Kavli Institute for Theoretical Physics, supported by NSF grant number PHY-1748958.

**Author contributions** The project is managed by A.W.-L. and P.V. The computing allocation proposals were written by E.B., P.V., A.W.-L., T.K., A.N.N. and E.R. The Feynman–Hellmann-theorem-inspired method was implemented by K.O., A.W.-L., C.C.C., C.B. and T.K. The lattice action was designed by C.J.M., A.W.-L. and K.O. The QUDA MDWF solver was optimized by M.A.C. The integration of the QUDA MDWF solver to Chroma was implemented by T.K. and B.J. The HMC generation of new ensembles was done with MILC software (v7.8.0) by E.R. and A.W.-L. The implementation of the code for mixed mesons using MILC and Chroma was done by E.R., B.J. and A.W.-L. The calculations were performed by E.R., A.W.-L., E.B., C.C.C., A.N.N., D.A.B. and H.M.-C., using the job management software written by E.B. The non-perturbative renormalization was performed by D.A.B., H.M.-C., N.G. and A.W.-L. The correlator analysis was performed by C.C.C., A.N.N., A.W.-L., C.B. and C.J.M. The extrapolation analysis was performed by C.C.C. and A.W.-L.

**Competing interests** The authors declare no competing interests.

#### Additional information

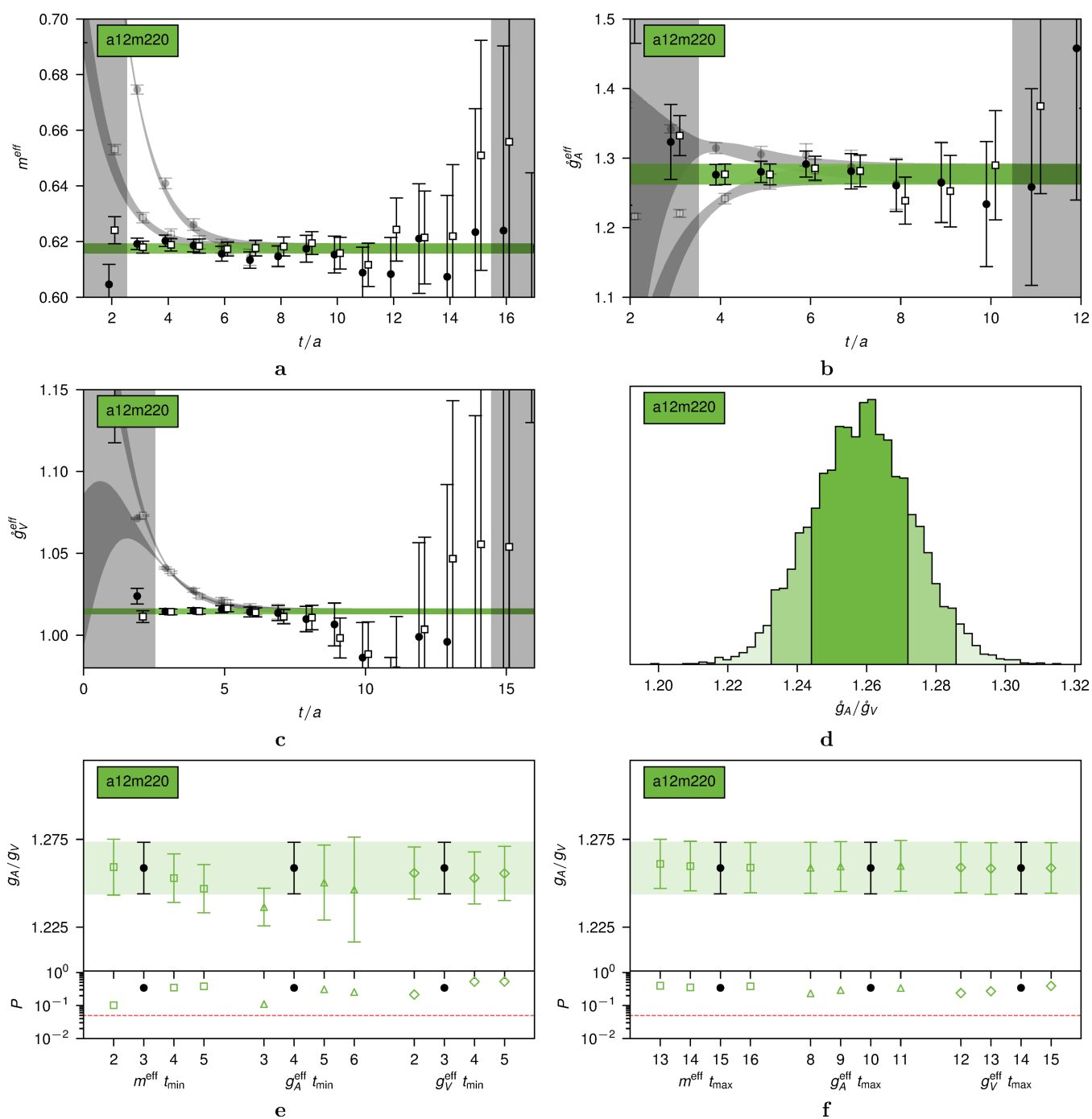
**Extended data** is available for this paper at <https://doi.org/10.1038/s41586-018-0161-8>.

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41586-018-0161-8>.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>.

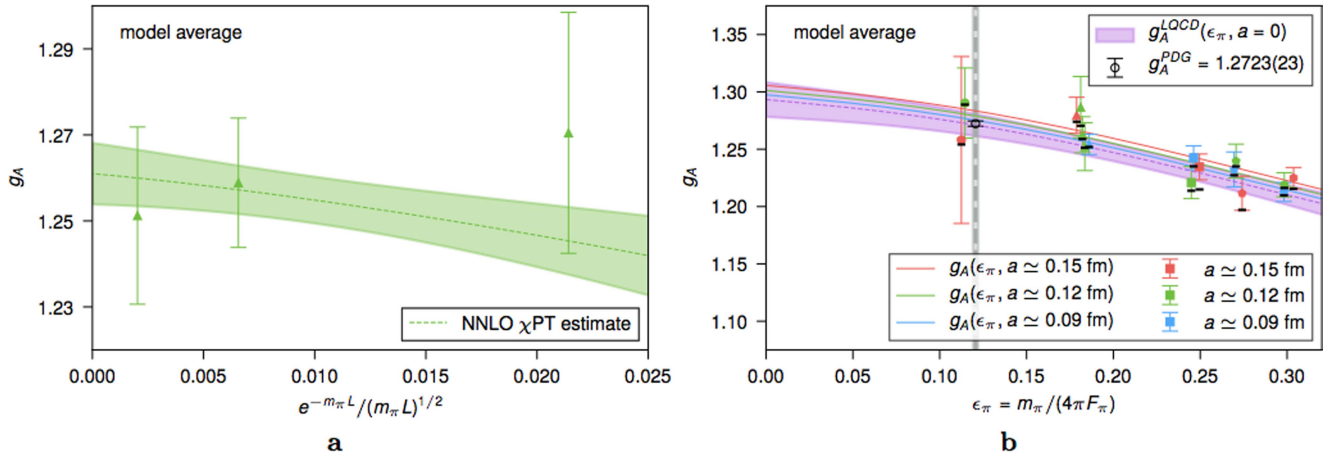
**Correspondence and requests for materials** should be addressed to A.W.-L.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



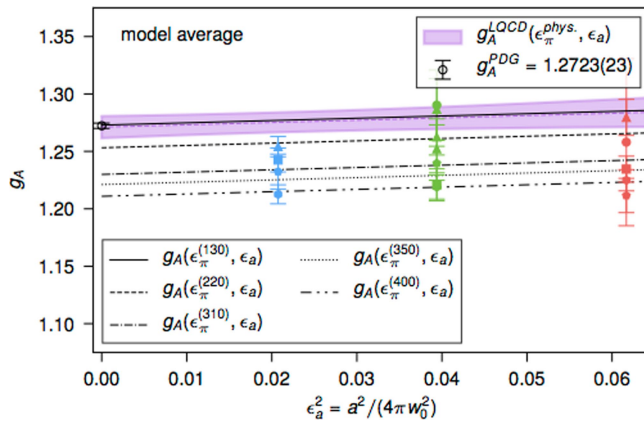
**Extended Data Fig. 1 | Correlator fit quality and stability.** **a–c,** Fit results for the effective mass ( $m^{\text{eff}}$ ), axial ( $g_A^{\text{eff}}$ ) and vector ( $g_V^{\text{eff}}$ ) Feynman–Hellmann ratios overlaid on top of correlator data. The black and white filled data points are ground state values determined by subtracting the excited state contributions from the raw correlation functions under bootstrap resampling. **d,** The distribution of  $g_A/g_V$  after 5,000 bootstrap resamplings. The green-shaded regions correspond to the 68% (dark

green) and 95% (light green) confidence intervals. All 5,000 bootstraps are shown, with no evidence of outliers. **e, f,** Stability of the correlation function analysis under varying  $t_{\min}$  and  $t_{\max}$  for  $m^{\text{eff}}$ ,  $g_A^{\text{eff}}$  and  $g_V^{\text{eff}}$ . The corresponding  $P$  values are shown in the bottom panel. The preferred simultaneous fit is highlighted by the solid black symbols. Uncertainties are one s.e.m.

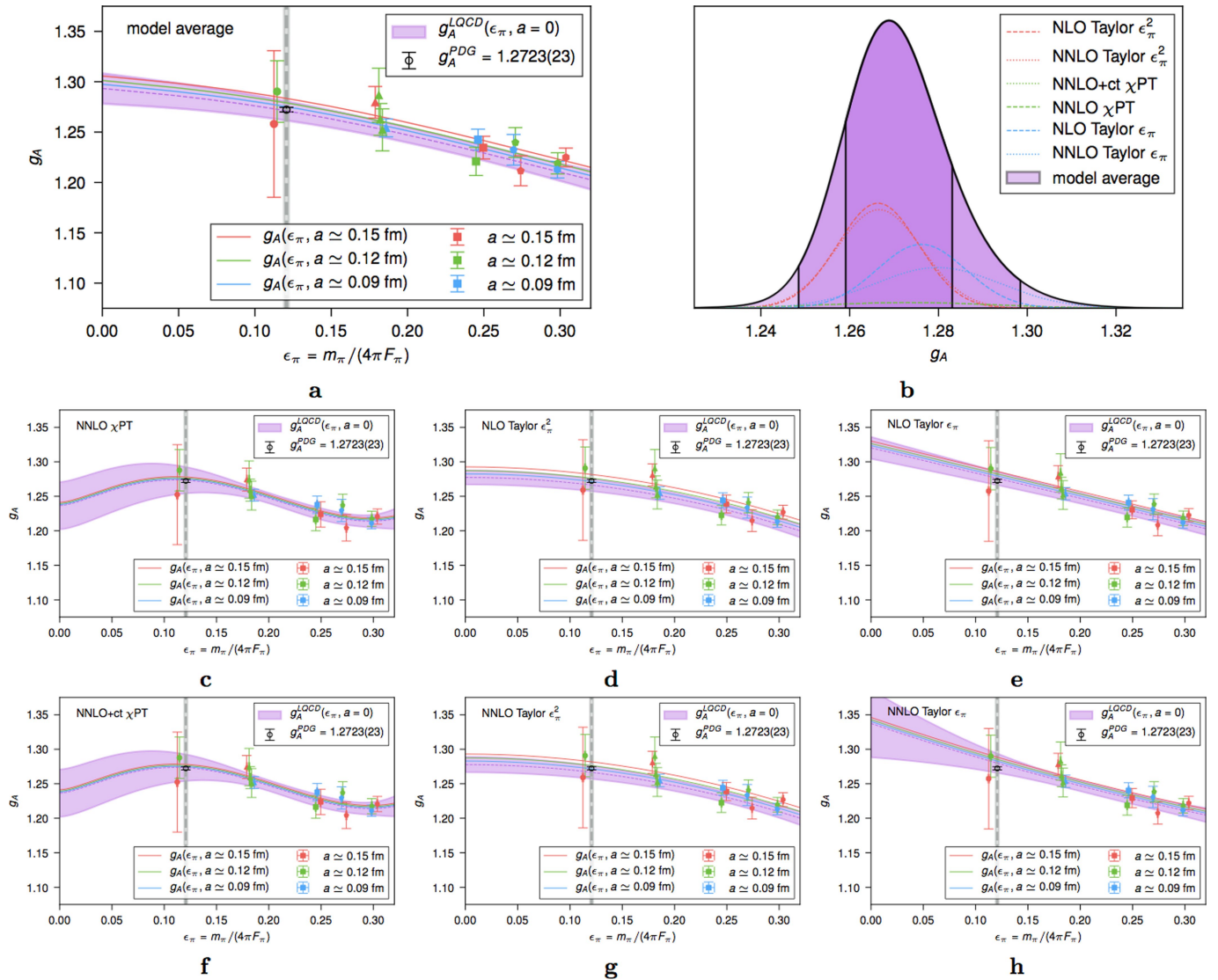


**Extended Data Fig. 2 | Infinite-volume extrapolation of  $g_A$ .** **a**, The three data points correspond to the  $a \approx 0.12$  fm and  $m_\pi \approx 220$  MeV ensembles with  $m_\pi L = \{5.36, 4.30, 3.25\}$ . The next-to-leading order (NLO) finite-volume dependence predicted from the model-averaged extrapolation (to all 16 data points) is shown by the green band, with the central value

indicated by the dashed green curve. **b**, Model-averaged extrapolation with finite-volume-adjusted data (coloured points). The central values of the raw data are denoted by a small black dash and, in all but one case, lie within one standard deviation of the finite-volume-adjusted result. Uncertainties are one s.e.m.

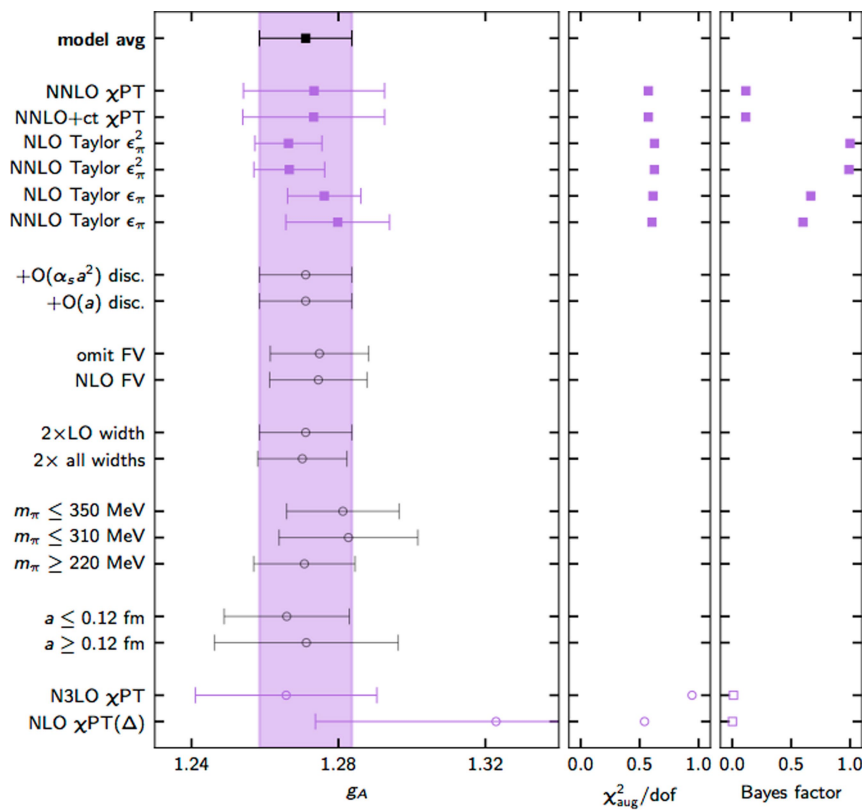


**Extended Data Fig. 3 | Continuum extrapolation of  $g_A$ .** The nucleon axial coupling as a function of  $\epsilon_a^2 = a^2 / (4\pi w_0^2)$ , where  $a$  is the lattice spacing and  $w_0$  is a hadronic length scale used to normalize LQCD calculations. The physical pion-mass limit is displayed by the magenta band, with the central value indicated by the dashed magenta curve. Additional curves with suppressed uncertainty bands are plotted for  $m_\pi \approx 130$  MeV (solid),  $m_\pi \approx 220$  MeV (dashed),  $m_\pi \approx 310$  MeV (dot-dashed),  $m_\pi \approx 350$  MeV (dotted) and  $m_\pi \approx 400$  MeV (dot-dot-dashed). Uncertainties are one s.e.m.



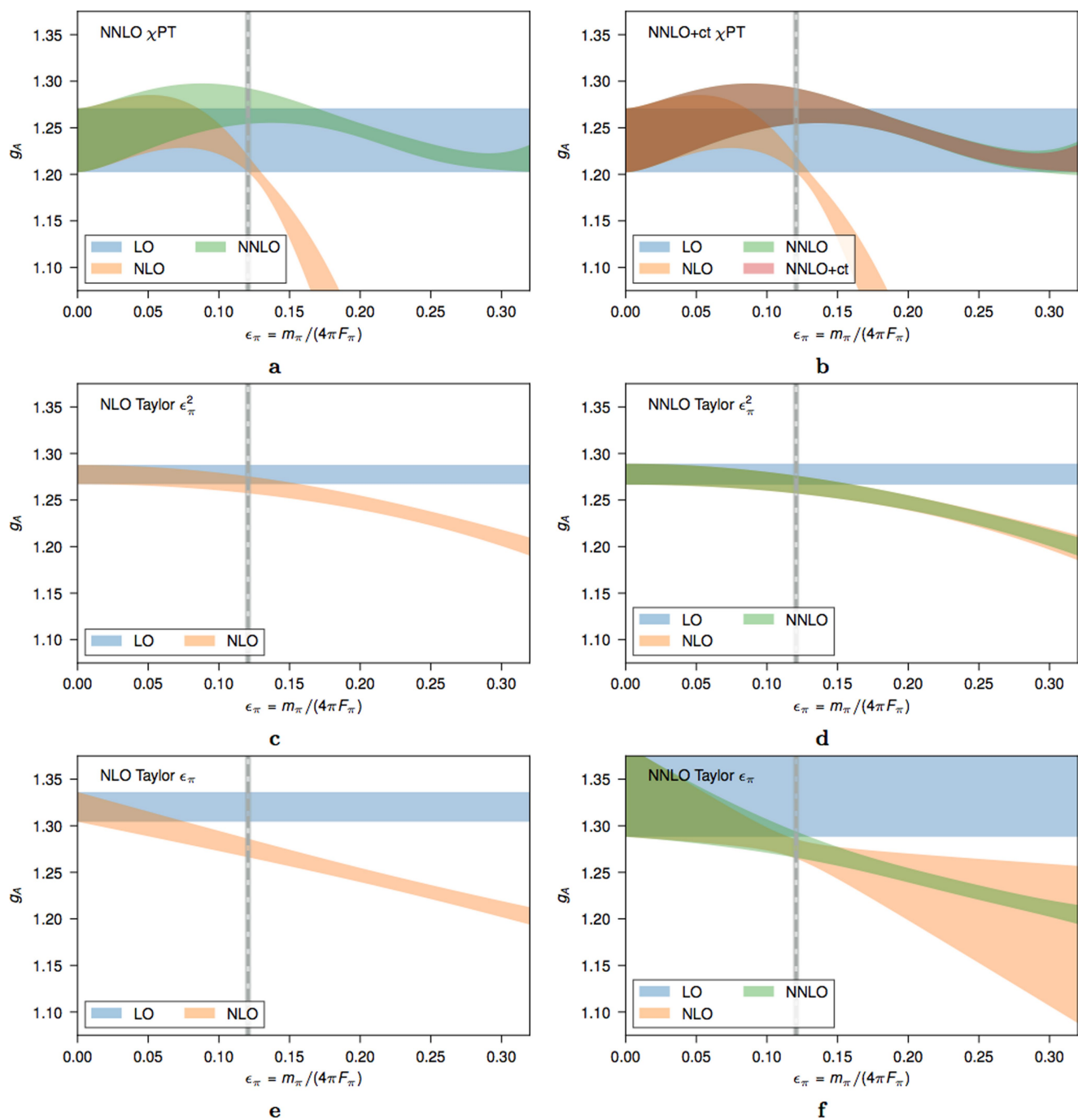
**Extended Data Fig. 4 | Model extrapolation plots. a**, Model-averaged extrapolation of  $g_A$  as a function of  $\epsilon_\pi$ , determined as described in Supplementary Information (section S.7A). **b**, Determination of  $g_A$  at the physical point from the model-averaging procedure. The magenta histogram is the final determination of  $g_A$ , constructed from a weighted average of the various models used in the extrapolation, which appear as the distributions lying inside the final histogram. **c–h**, The resulting

extrapolation of  $g_A$  as a function of  $\epsilon_\pi$  for each of the six models used in the averaging procedure (see Supplementary Information, section S.6). The magenta band is the resulting 68% confidence interval of the continuum, infinite-volume extrapolated value of  $g_A$  as a function of  $\epsilon_\pi$ . The red, green and blue curves are the central values of  $g_A$  versus  $\epsilon_\pi$  at fixed lattice spacings of 0.15 fm, 0.12 fm and 0.09 fm, respectively. Uncertainties are one s.e.m.



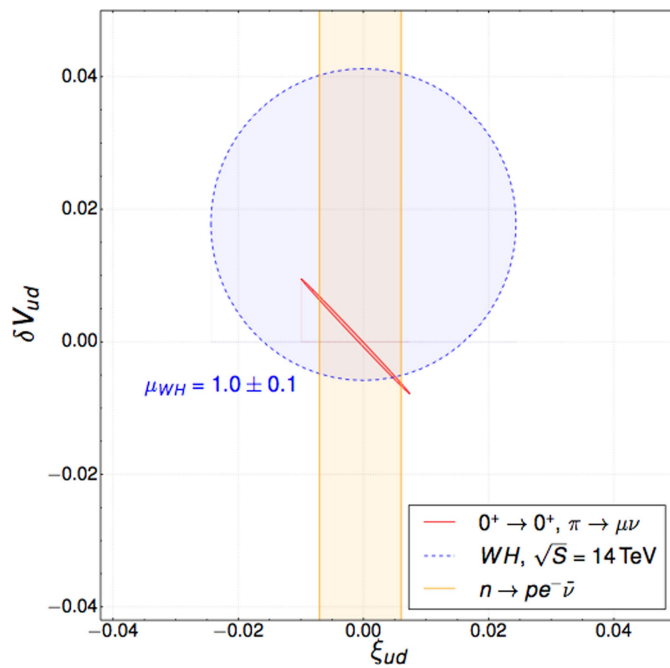
**Extended Data Fig. 5 | Stability and convergence of the chiral-continuum extrapolation.** In the left panel, the model-averaged result ('model avg') is the black square. The vertical magenta band is the resulting 68% confidence band. The next six values are results from individual extrapolations that go into the model average, described in Supplementary Information, section S.7A. Uncertainties are one s.e.m. 'ct', counter-term; 'FV', finite volume; 'disc.', discretization;  $\alpha_s = g^2/(4\pi)$ , where  $g$  is the quark-gluon coupling of QCD. The middle panel shows the augmented  $\chi^2$  ( $\chi_{\text{aug}}^2$ ) per degree of freedom (dof), where  $\chi_{\text{aug}}^2$  is the sum of the  $\chi^2$  values from the data and from the priors. All fits have 16 degrees of freedom because each prior is counted as a data point. The right panel shows the resulting Bayes factors normalized by the NLO Taylor  $\epsilon_\pi^2$  Bayes factor, which is found to be the largest among them. These normalized Bayes factors are used as relative weights in the model-averaging procedure. The stability of the extrapolation analysis is tested by including

additional discretization terms, omitting the predicted NLO finite-volume corrections, increasing the prior widths on the leading order (LO) and all low-energy constants, and applying cuts on the pion masses considered and on the discretization scales included. All variations are contained within  $1\sigma$  of the model-average value, with most being substantially smaller than  $1\sigma$  from the central value. Finally, we show the resulting extrapolation from the complete next-to-next-to-next-to-leading order (N3LO) chiral perturbation theory analysis and from the NLO chiral perturbation theory analysis with  $\Delta$  degrees of freedom ( $\chi\text{PT}(\Delta)$ ). The N3LO fit is not included in the average because it has five unknown low-energy constants and we have only five different pion mass values. The NLO  $\chi\text{PT}(\Delta)$  value is not included because it requires input from phenomenology and is thus not a pure lattice QCD prediction, and also the next-to-next-to-leading order (NNLO)  $\chi\text{PT}(\Delta)$  extrapolation function is not known, so a test of stability and convergence is not possible.



**Extended Data Fig. 6 | Convergence of  $g_A$ .** a–f, Order-by-order contribution to the extrapolation of  $g_A$  for the six different models that enter in the final model-averaged result (see Supplementary Information, section S.6). The low-energy constants are determined by the full fit from

each model. Higher orders are added successively, producing the final reconstruction of the extrapolation when all contributions up to a given order are included.



**Extended Data Fig. 7 | Constraint on right-handed beyond-standard-model currents.** Measurements of cold neutron decays ( $n \rightarrow pe^- \bar{\nu}$ ;  $n$ , neutron;  $p$ , proton;  $e^-$ , electron;  $\bar{\nu}$ , antineutrino) provide some of the most stringent constraints on new physics. A recent comparison of constraints from low-energy experiments and colliders found comparable constraints on right-handed beyond-standard-model currents<sup>29</sup>. The figure has been adapted from figure 12 of ref. <sup>29</sup> using our determination of  $g_A$ . The vertical orange band is the constraint on the right-handed coupling ( $\xi_{ud}$ ) from our result. The blue circle arises from collider constraints on  $W$ - and Higgs-boson production (WH) at collision energy  $\sqrt{S} = 14$  TeV, and the diagonal red band is from pion decays (long direction;  $\pi \rightarrow \mu \bar{\nu}$ , where  $\mu$  is a muon) and super-allowed  $0^+ \rightarrow 0^+$  nuclear decays, which constrain corrections to the axial (left ( $\delta V_{ud}$ ) minus right) and vector (left plus right) currents, respectively.

Extended Data Table 1 | Data and inputs for the chiral–continuum extrapolation

ensemble	$\epsilon_\pi$	$m_\pi L$	$a/w_0$	$\alpha_S$	$g_A$
a15m400	0.30374(53)	4.8451(49)	0.8804(3)	0.58801	1.216(06)
a15m350	0.27411(50)	4.2359(47)	0.8804(3)	0.58801	1.198(13)
a15m310	0.24957(36)	3.7772(48)	0.8804(3)	0.58801	1.215(12)
a15m220	0.18084(30)	3.9673(45)	0.8804(3)	0.58801	1.274(14)
a15m130	0.11340(74)	3.227(19)	0.8804(3)	0.58801	1.270(72)
a12m400	0.29841(52)	5.8428(39)	0.7036(5)	0.53796	1.217(10)
a12m350	0.27063(69)	5.1352(49)	0.7036(5)	0.53796	1.236(14)
a12m310	0.24485(50)	4.5282(41)	0.7036(5)	0.53796	1.214(13)
a12m220S	0.18419(57)	3.2523(76)	0.7036(5)	0.53796	1.272(28)
a12m220	0.18221(42)	4.2959(56)	0.7036(5)	0.53796	1.259(15)
a12m220L	0.18156(44)	5.3604(61)	0.7036(5)	0.53796	1.252(21)
a12m130	0.11347(50)	3.899(12)	0.7036(5)	0.53796	1.292(30)
a09m400	0.29818(53)	5.7965(46)	0.5105(3)	0.43356	1.210(08)
a09m350	0.26949(57)	5.0502(62)	0.5105(3)	0.43356	1.228(15)
a09m310	0.24619(44)	4.5035(38)	0.5105(3)	0.43356	1.236(11)
a09m220	0.18197(37)	4.6990(32)	0.5105(3)	0.43356	1.253(09)

$\epsilon_\pi$ ,  $m_\pi L$  and renormalized values of  $g_A$  determined in this work. The lattice spacing  $a/w_0$  and strong coupling constant  $\alpha_S$  are obtained as described previously<sup>21</sup>. We use the  $a/w_0$  value determined at the physical pion mass for each lattice spacing. The quantities  $\epsilon_\pi$ ,  $a/w_0$  and  $m_\pi L$  are used to guide the chiral, continuum and infinite-volume extrapolations, respectively.

**Extended Data Table 2 | Highly improved staggered quark gauge configurations and valence sector parameters**

HISQ gauge configuration parameters							valence parameters							
abbr.	$N_{\text{cfg}}$	volume	$\sim a$ [fm]	$m_l/m_s$	$\sim m_{\pi_5}$ [MeV]	$\sim m_{\pi_5} L$	$N_{\text{src}}$	$L_5/a$	$aM_5$	$b_5$	$c_5$	$am_l^{\text{val.}}$	$\sigma_{\text{smr}}$	$N_{\text{smr}}$
a15m400	1000	$16^3 \times 48$	0.15	0.334	400	4.8	8	12	1.3	1.5	0.5	0.0278	3.0	30
a15m350	1000	$16^3 \times 48$	0.15	0.255	350	4.2	16	12	1.3	1.5	0.5	0.0206	3.0	30
a15m310	1960	$16^3 \times 48$	0.15	0.2	310	3.8	24	12	1.3	1.5	0.5	0.01580	4.2	60
a15m220	1000	$24^3 \times 48$	0.15	0.1	220	4.0	24	16	1.3	1.75	0.75	0.00712	4.5	60
a15m130	1000	$32^3 \times 48$	0.15	0.036	130	3.2	5	24	1.3	2.25	1.25	0.00216	4.5	60
a12m400	1000	$24^3 \times 64$	0.12	0.334	400	5.8	8	8	1.2	1.25	0.25	0.02190	3.0	30
a12m350	1000	$24^3 \times 64$	0.12	0.255	350	5.1	8	8	1.2	1.25	0.25	0.01660	3.0	30
a12m310	1053	$24^3 \times 64$	0.12	0.2	310	4.5	8	8	1.2	1.25	0.25	0.01260	3.0	30
a12m220S	1000	$24^3 \times 64$	0.12	0.1	220	3.2	4	12	1.2	1.5	0.5	0.00600	6.0	90
a12m220	1000	$32^3 \times 64$	0.12	0.1	220	4.3	4	12	1.2	1.5	0.5	0.00600	6.0	90
a12m220L	1000	$40^3 \times 64$	0.12	0.1	220	5.4	4	12	1.2	1.5	0.5	0.00600	6.0	90
a12m130	1000	$48^3 \times 64$	0.12	0.036	130	3.9	3	20	1.2	2.0	1.0	0.00195	7.0	150
a09m400	1201	$32^3 \times 64$	0.09	0.335	400	5.8	8	6	1.1	1.25	0.25	0.0160	3.5	45
a09m350	1201	$32^3 \times 64$	0.09	0.255	350	5.1	8	6	1.1	1.25	0.25	0.0121	3.5	45
a09m310	784	$32^3 \times 96$	0.09	0.2	310	4.5	8	6	1.1	1.25	0.25	0.00951	7.5	167
a09m220	1001	$48^3 \times 96$	0.09	0.1	220	4.7	6	8	1.1	1.25	0.25	0.00449	8.0	150

The highly improved staggered quark (HISQ) ensembles used in this work. In the abbreviated naming convention<sup>2</sup> ('abbr.'): 'a15m310' stands for an ensemble with  $a \approx 0.15$  fm and  $m_\pi \approx 310$  MeV. The table also shows the number of configurations ( $N_{\text{cfg}}$ ), lattice volume, approximate lattice spacing ( $a$ ), ratio of the input light and strange sea-quark masses ( $m_l/m_s$ ), approximate HISQ taste-5 pion mass ( $m_{\pi_5}$ ) and approximate value of  $m_{\pi_5} L$ . The values were obtained from table 1 of ref. <sup>30</sup> with increased number of configurations. With the HISQ gauge configurations, we generate Möbius domain-wall propagators at a number of sources per configuration, with the fifth-dimensional extent  $L_5/a$  such that the residual chiral symmetry breaking quantity,  $m_{\text{res}}$ , is minimized at the domain-wall mass  $aM_5$ , with the Möbius kernel defined by the parameters  $b_5$  and  $c_5$ , and valence light-quark masses  $am_l^{\text{val.}}$ . We also list the width  $\sigma_{\text{smr}}$  and iteration count  $N_{\text{smr}}$  of the SHELL\_SOURCE and the GAUGE\_INV\_GAUSSIAN smearing algorithm in Chroma.

# Observation of anisotropic magneto–Peltier effect in nickel

Ken-ichi Uchida<sup>1,2,3,4\*</sup>, Shunsuke Daimon<sup>4,5,7</sup>, Ryo Iguchi<sup>1</sup> & Eiji Saitoh<sup>3,4,5,6,7</sup>

**The Peltier effect, discovered in 1834, converts a charge current into a heat current in a conductor, and its performance is described by the Peltier coefficient, which is defined as the ratio of the generated heat current to the applied charge current<sup>1,2</sup>. To exploit the Peltier effect for thermoelectric cooling or heating, junctions of two conductors with different Peltier coefficients have been believed to be indispensable. Here we challenge this conventional wisdom by demonstrating Peltier cooling and heating in a single material without junctions. This is realized through an anisotropic magneto–Peltier effect in which the Peltier coefficient depends on the angle between the directions of a charge current and magnetization in a ferromagnet. By using active thermography techniques<sup>3–10</sup>, we observe the temperature change induced by this effect in a plain nickel slab. We find that the thermoelectric properties of the ferromagnet can be redesigned simply by changing the configurations of the charge current and magnetization, for instance, by shaping the ferromagnet so that the current must flow around a curve. Our experimental results demonstrate the suitability of nickel for the anisotropic magneto–Peltier effect and the importance of spin–orbit interaction in its mechanism. The anisotropic magneto–Peltier effect observed here is the missing thermoelectric phenomenon in ferromagnetic materials—the Onsager reciprocal of the anisotropic magneto–Seebeck effect previously observed in ferromagnets—and its simplicity might prove useful in developing thermal management technologies for electronic and spintronic devices.**

Electron transport phenomena in conductors are phenomenologically described by the electrical conductivity, thermal conductivity and thermoelectric coefficients<sup>1,2</sup>. In ferromagnetic materials, the coefficients are known to be dependent on the direction of the magnetization because of the concerted action of spin-polarized electron transport and spin–orbit interaction, inducing a variety of electric, thermal and thermoelectric transport phenomena<sup>11–13</sup>. A typical example is the anisotropic magnetoresistance (AMR) in ferromagnets<sup>14–16</sup>, where the electrical resistivity  $\rho$  depends on the angle  $\theta$  between the directions of the charge current  $J_c$  and the magnetization  $M$ ; in isotropic ferromagnets, the AMR obeys the following equation:

$$\rho(\theta) = \rho_{\perp} + (\rho_{\parallel} - \rho_{\perp}) \cos^2 \theta \quad (1)$$

where  $\rho_{\perp(\parallel)}$  is the resistivity for  $M \perp J_c$  ( $M \parallel J_c$ ). In a similar manner to the AMR, thermoelectric transport phenomena in ferromagnets can also be affected by the spin–orbit interaction. In fact, the Seebeck effect (in which a temperature gradient produces a voltage) in ferromagnets is dependent on the direction of  $M$ , and this is called the anisotropic magneto–thermopower or anisotropic magneto–Seebeck effect (AMSE)<sup>12,17–25</sup>. Its reciprocal process remain to be observed: the anisotropic magneto–Peltier effect (AMPE) is the missing piece of thermoelectric phenomena in ferromagnetic materials.

Here we report the observation of the AMPE in a ferromagnetic metal and, by making use of the AMPE, demonstrate Peltier cooling/

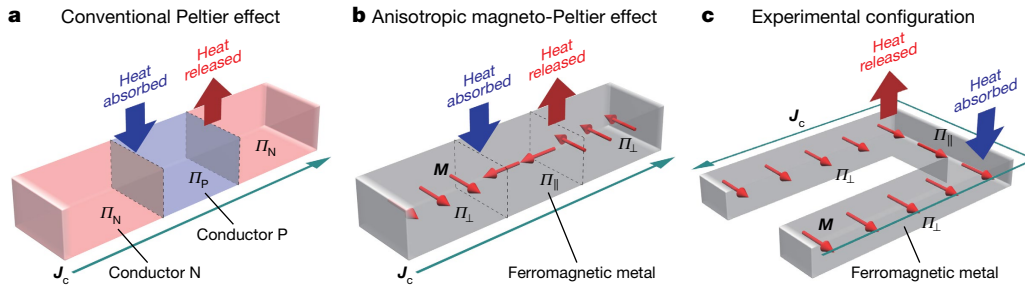
heating in a single material without junctions, which would be impossible through the conventional Peltier effect alone (Fig. 1a, b). The Peltier effect is driven by the entropy current that accompanies  $J_c$ , and its efficiency for simple metals depends on the density of states, group velocity and relaxation time of electrons near the Fermi energy in the semiclassical approximation<sup>1</sup>. In ferromagnets, these parameters can be dependent on the angle between the entropy current ( $\parallel J_c$ ) and the direction of  $M$  in the presence of the spin–orbit interaction, resulting in the finite  $\theta$  dependence of the Peltier coefficient  $\Pi$ : this is the AMPE. Judging from the symmetry of the AMSE<sup>12</sup>, the  $\theta$  dependence of  $\Pi$  in isotropic ferromagnets is expected to be similar to equation (1):

$$\Pi(\theta) = \Pi_{\perp} + (\Pi_{\parallel} - \Pi_{\perp}) \cos^2 \theta \quad (2)$$

where  $\Pi_{\perp(\parallel)}$  is the Peltier coefficient for  $M \perp J_c$  ( $M \parallel J_c$ ). Therefore, when  $\Pi_{\parallel} - \Pi_{\perp} \neq 0$ , by forming a non-uniform magnetization configuration along  $J_c$  in a ferromagnet, Peltier cooling/heating is generated between areas with different  $\theta$  values even in the absence of junction structures; the non-uniform magnetization configuration works as virtual Peltier junctions (Fig. 1a, b). To observe the temperature change induced by the AMPE, we construct a U-shaped ferromagnetic metal with uniform magnetization and apply a charge current along the U-shaped structure (Fig. 1c). This is equivalent to the configuration shown in Fig. 1b, giving rise to heat absorption and release at the corners of the U-shaped structure due to the AMPE.

Direct observation of the AMPE is realized by means of active infrared emission microscopy called lock-in thermography (LIT)<sup>3–10</sup>, which makes it possible to visualize the spatial distribution of temperature modulation induced by thermoelectric effects with high temperature and spatial resolutions<sup>6,9</sup>. In the LIT measurements, when a periodic charge current is applied to a sample, thermal images oscillating with the same frequency as the current can be extracted through Fourier analysis (Fig. 2a). The obtained thermal images are transformed into images of the lock-in amplitude  $A$  and phase  $\phi$ . The  $A$  image shows the distribution of the magnitude of the current-induced temperature modulation, and the  $\phi$  image shows the distribution of the sign of the temperature modulation, as well as the time delay due to thermal diffusion, where the  $A$  and  $\phi$  values are defined in the ranges of  $A \geq 0$  and  $0^\circ \leq \phi < 360^\circ$ , respectively. To observe the AMPE using the LIT method, we measured the spatial distribution of infrared radiation thermally emitted from the surface of a U-shaped polycrystalline Ni slab by using an infrared camera while applying a rectangularly modulated alternating charge current with amplitude  $J_c$ , frequency  $f$  and zero direct current (d.c.) offset to the slab (see Methods). By extracting the first harmonic response of detected thermal images, we can separate the contribution of thermoelectric effects ( $\propto J_c$ ) from that of Joule heating ( $\propto J_c^2$ ), because the Joule heating generated by such a rectangular alternating current is constant in time<sup>6,8,9</sup>. The maximum  $J_c$  value applied to the Ni slab is 1.00 A, which corresponds to a charge-current density of  $1 \times 10^7$  A m<sup>−2</sup>. Because this value is several orders of magnitude smaller than typical threshold currents for

<sup>1</sup>National Institute for Materials Science, Tsukuba, Japan. <sup>2</sup>PRESTO, Japan Science and Technology Agency, Saitama, Japan. <sup>3</sup>Center for Spintronics Research Network, Tohoku University, Sendai, Japan. <sup>4</sup>Institute for Materials Research, Tohoku University, Sendai, Japan. <sup>5</sup>Advanced Institute for Materials Research, Tohoku University, Sendai, Japan. <sup>6</sup>Advanced Science Research Center, Japan Atomic Energy Agency, Tokai, Japan. <sup>7</sup>Present address: Department of Applied Physics, The University of Tokyo, Tokyo, Japan. \*e-mail: UCHIDA.Kenichi@nims.go.jp



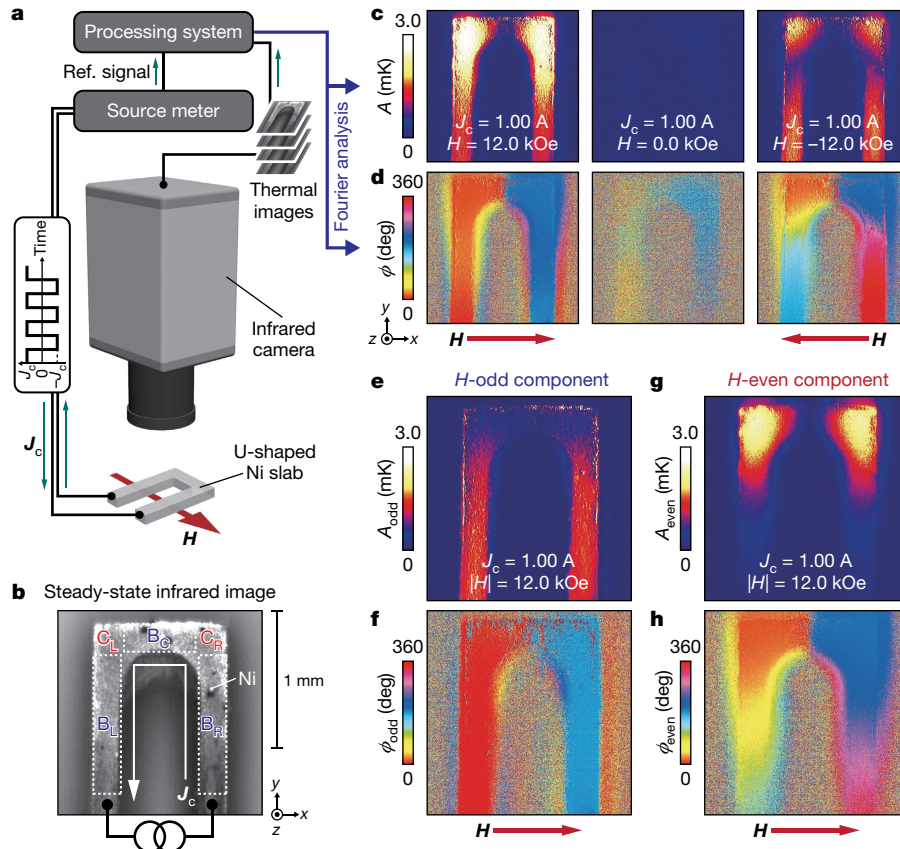
**Fig. 1 | Conventional Peltier and anisotropic magneto-Peltier effects.** **a**, Schematic of the conventional Peltier effect in a junction comprising two conductors P and N. When a charge current  $J_c$  is applied to the junction, the Peltier effect induces heat absorption or release at the P/N interface owing to the difference between the Peltier coefficient of P,  $\Pi_P$ , and that of N,  $\Pi_N$ . **b**, Schematic of the anisotropic magneto-Peltier effect

(AMPE) in a ferromagnetic metal. When  $J_c$  is applied to the ferromagnet with the magnetization vector  $M$ , the AMPE induces heat absorption or release even in the absence of junctions, because of the difference between the Peltier coefficient for the region with  $M \perp J_c$ ,  $\Pi_{\perp}$ , and that with  $M \parallel J_c$ ,  $\Pi_{\parallel}$ . **c**, Experimental configuration for measuring the AMPE in a U-shaped ferromagnet.

inducing domain-wall motions<sup>26,27</sup>, the thermal responses reported in this study are irrelevant to such magnetization dynamics. In general, LIT measurements with higher  $f$  values make it easier to identify heat-source positions because the temperature broadening due to thermal diffusion is reduced by increasing  $f$ . Therefore, we set  $f = 25.0$  Hz, the maximum lock-in frequency of our system, except for the  $f$ -dependent measurements shown in Extended Data Fig. 1. The detected infrared radiation is converted into temperature information through the calibration detailed in Methods. To enhance infrared emissivity and

to ensure uniform emission properties, the surface of the Ni slab was coated with insulating black ink, which has emissivity  $>0.95$ . During the LIT measurements, an in-plane magnetic field  $H$  (with magnitude  $H$ ) was applied along the  $x$  direction, except for the measurements of  $H$ -angle dependence. When  $|H| > 5$  kOe, the  $M$  direction of the Ni slab is aligned along the  $H$  direction. All the measurements were carried out at room temperature and atmospheric pressure.

For the pure detection of the AMPE, it is important to distinguish the AMPE signals from other thermoelectric effects. First, the contribution



**Fig. 2 | Thermal imaging of anisotropic magneto-Peltier effect.** **a**, Measurements of the AMPE in a U-shaped Ni slab using the LIT method.  $H$  and  $J_c$  denote the external magnetic field vector with magnitude  $H$ , generated by an electromagnet, and the rectangularly modulated charge current with amplitude  $J_c$  and frequency  $f$ , respectively. **b**, Steady-state infrared image of the U-shaped Ni slab with black-ink coating at thermal equilibrium. The areas labelled  $B_{L/C/R}$ , surrounded by white dotted lines, and the areas  $C_{L/R}$ , the corners of the U-shaped

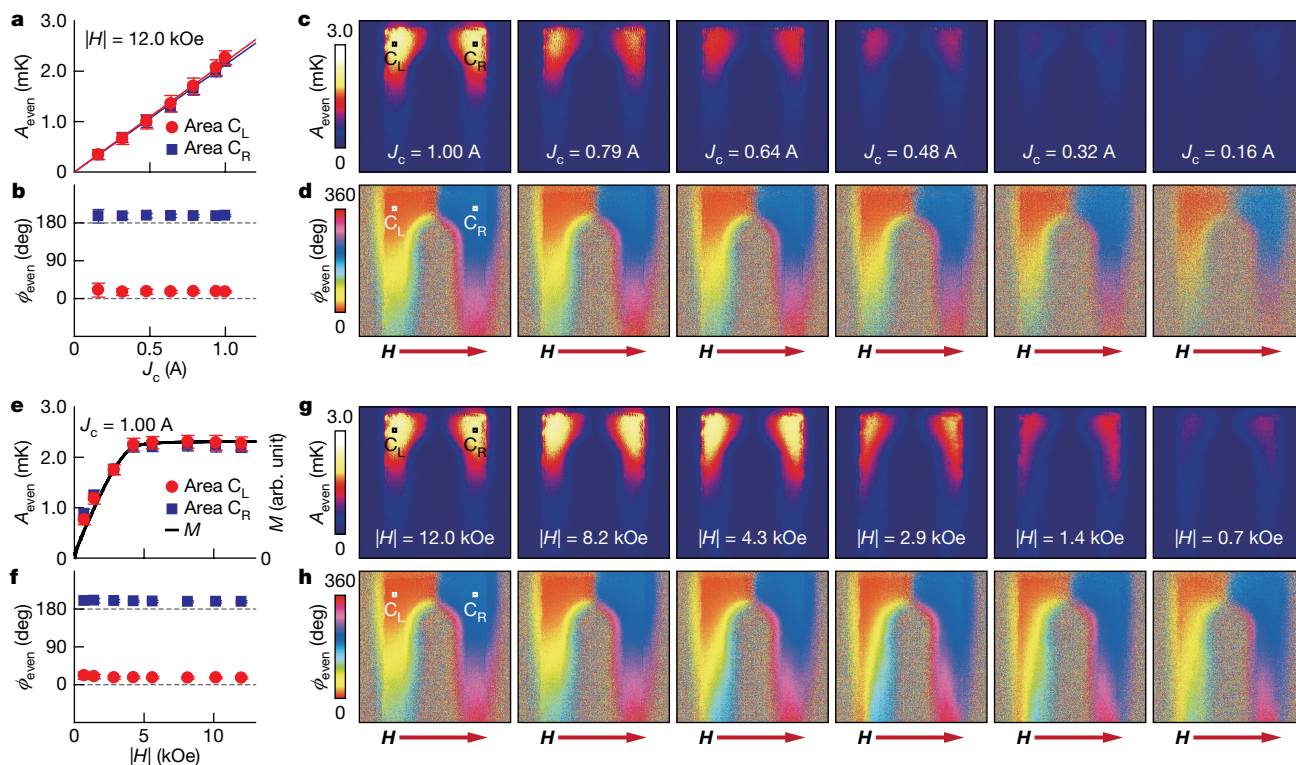
structure, are the parts of the Ni slab. **c, d**, Lock-in amplitude  $A$  and phase  $\phi$  images for the U-shaped Ni slab at  $J_c = 1.00$  A. The left, centre and right images show the results at  $H = +12.0$ ,  $0.0$  and  $-12.0$  kOe, respectively. **e, f**,  $A_{\text{odd}}$  and  $\phi_{\text{odd}}$  images for the U-shaped Ni slab at  $J_c = 1.00$  A and  $|H| = 12.0$  kOe. **g, h**,  $A_{\text{even}}$  and  $\phi_{\text{even}}$  images for the U-shaped Ni slab at  $J_c = 1.00$  A and  $|H| = 12.0$  kOe.  $A_{\text{odd( even)}}$  and  $\phi_{\text{odd( even)}}$  are the  $H$ -odd ( $H$ -even) components of the lock-in amplitude and phase, respectively.

coming from the conventional  $H$ -independent Peltier effect is excluded by measuring the  $H$  dependence of the LIT images (we note that the temperature modulation due to the conventional Peltier effect does not appear in our Ni sample around the U-shaped structure because of the absence of junctions, except for parts connected to external wires far from the viewing area). Second, in the present set-up, we have to separate the anomalous Ettingshausen effect (AEE)<sup>12,28</sup>, where a heat current is generated in the direction of the cross-product of  $\mathbf{M}$  and  $\mathbf{J}_c$ , from the AMPE. These can be distinguished from each other by considering their symmetries; in the set-up shown in Fig. 2b, the AMPE signal is expected to appear at the corners  $C_{L/R}$  of the U-shaped structure, that is, the region between the area  $B_{L/R}$  satisfying  $\mathbf{M} \perp \mathbf{J}_c$  and the area  $B_C$  satisfying  $\mathbf{M} \parallel \mathbf{J}_c$  (Figs. 1c and 2b), whereas the AEE signal appears in  $B_{L/R}$ . According to equation (2), the AMPE exhibits an even dependence on the  $\mathbf{M}$  direction, whereas the AEE has an odd dependence<sup>12,28</sup>. Therefore, the AMPE and AEE contributions can also be distinguished by measuring the  $H$  dependence of the LIT images.

Figure 2c, d shows the  $A$  and  $\phi$  images for the U-shaped Ni slab at  $J_c = 1.00$  A. When a magnetic field of  $H = +12.0$  kOe, much greater than the saturation field for Ni, is applied to the sample along the  $x$  direction, a clear current-induced temperature modulation appears (left images of Fig. 2c, d). This temperature modulation is not due to the conventional Peltier effect, as no signal is generated in the absence of the magnetic field (centre images of Fig. 2c, d); without the magnetic field, the Ni slab has no net magnetization because there is negligible coercive force (see the magnetization curve in Fig. 3e and note that small Oersted fields induced by the applied currents ( $<3$  Oe) do not affect the magnetization of the Ni slab). Surprisingly, the distribution of the current-induced temperature modulation strongly depends on the sign of  $H$ ; at  $H = -12.0$  kOe, a complicated pattern appears (the right images of Fig. 2c, d). We found that the signal consists of two

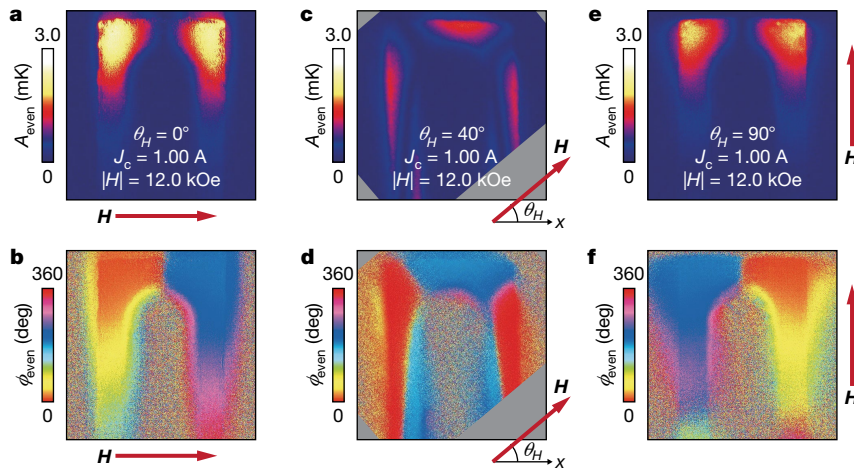
contributions showing the odd and even dependence on the  $H$  sign. To separate these contributions, we extract the  $H$ -odd ( $H$ -even) component from the raw LIT images, as shown in Fig. 2e, f (2g, h), in which the LIT amplitude and phase of the  $H$ -odd ( $H$ -even) component are denoted by  $A_{\text{odd(even)}}$  and  $\phi_{\text{odd(even)}}$ , respectively. The signal with  $H$ -odd dependence is generated only in the areas  $B_{L/R}$  of the U-shaped structure, where  $\mathbf{M} \perp \mathbf{J}_c$ , which is attributed to the heat current generated by the AEE along the  $z$  direction (Fig. 2e, f and Extended Data Fig. 2). By contrast, the signal with  $H$ -even dependence is generated around the corners of the U-shaped structure (Fig. 2g), and the sign of the temperature modulation at  $C_L$  is opposite to that at  $C_R$  (Fig. 2h) because the  $\phi_{\text{even}}$  difference between  $C_L$  and  $C_R$  is approximately  $180^\circ$ . The  $\phi_{\text{even}}$  values show that the input charge current and output temperature modulation oscillate with the same (opposite) phase at  $C_L$  ( $C_R$ ), indicating that heat is released (absorbed) at  $C_L$  ( $C_R$ ). This is the behaviour expected for the AMPE (Fig. 1c and equation (2)).

To clarify the origin of the temperature modulation with the  $H$ -even dependence, we performed systematic measurements. As shown in Fig. 3a–d, the  $A_{\text{even}}$  value at  $C_{L/R}$  is proportional to  $J_c$ , whereas the  $\phi_{\text{even}}$  difference between  $C_L$  and  $C_R$  at finite  $J_c$  values remains at about  $180^\circ$ . This result indicates that the temperature modulation is due to thermoelectric effects, appearing in linear response to the charge current applied to the Ni slab. Figure 3e–h shows the  $|H|$  dependence of  $A_{\text{even}}$  and  $\phi_{\text{even}}$  for the same sample, confirming that the magnitude of the temperature modulation gradually decreases with decreasing  $|H|$  in response to the magnetization process of the Ni slab (Fig. 3e). These behaviours are also consistent with the features of the AMPE mentioned above; the temperature modulation at  $C_{L/R}$  originates from the difference in the Peltier coefficient between the area  $B_{L/R}$  possessing  $\Pi_\perp$  and the area  $B_C$  possessing  $\Pi_\parallel$  (equation (2)). Here, the sign of the observed AMPE signals indicates  $\Pi_\parallel > \Pi_\perp$ . By combining the



**Fig. 3 | Dependence on charge current and on magnetic field.** **a, b,**  $J_c$  dependence of  $A_{\text{even}}$  and  $\phi_{\text{even}}$  on the areas  $C_{L/R}$  of the U-shaped Ni slab at  $|H| = 12.0$  kOe (red circles and blue squares). The solid lines in **a** represent the linear fits to the data. **c, d,**  $A_{\text{even}}$  and  $\phi_{\text{even}}$  images for the U-shaped Ni slab at  $|H| = 12.0$  kOe for various values of  $J_c$ . **e, f,**  $|H|$  dependence of  $A_{\text{even}}$  and  $\phi_{\text{even}}$  on  $C_{L/R}$  at  $J_c = 1.00$  A (red circles and blue squares) and the magnetization  $M$  curve of the Ni slab (black line). The  $M$ - $H$  curve

was measured by vibrating sample magnetometry while applying  $H$  to the U-shaped Ni slab along the  $x$  direction. **g, h,**  $A_{\text{even}}$  and  $\phi_{\text{even}}$  images at  $J_c = 1.00$  A for various values of  $|H|$ . The data points in **a** and **b** (**e** and **f**) are respectively obtained by averaging the  $A_{\text{even}}$  and  $\phi_{\text{even}}$  values on the areas defined by the squares with the size  $11 \times 11$  pixels in the leftmost images of **c** and **d** (**g** and **h**). The error bars represent the standard deviation of the data in the corresponding squares.



**Fig. 4 | Dependence on magnetic field angle.** **a, b,**  $A_{\text{even}}$  and  $\phi_{\text{even}}$  images for the U-shaped Ni slab at the magnetic field angle of  $\theta_H = 0^\circ$ ,  $J_c = 1.00$  A and  $|H| = 12.0$  kOe. **c, d,**  $A_{\text{even}}$  and  $\phi_{\text{even}}$  images at  $\theta_H = 40^\circ$ ,  $J_c = 1.00$  A and  $|H| = 12.0$  kOe. **e, f,**  $A_{\text{even}}$  and  $\phi_{\text{even}}$  images at  $\theta_H = 90^\circ$ ,  $J_c = 1.00$  A

experimental results with numerical calculations based on finite element methods, we estimate the anisotropy of the Peltier coefficient for Ni to be  $\Pi_{\parallel} - \Pi_{\perp} = 0.11 \times 10^{-3}$  V, the magnitude of which is about 2% of the conventional Peltier coefficient of Ni,  $\Pi = -6 \times 10^{-3}$  V (ref. <sup>21</sup>), at room temperature (Extended Data Fig. 3 and Methods). This result is consistent with the measurements of the AMSE in Ni, where the anisotropy of the Seebeck coefficient is estimated to be 3% (Extended Data Fig. 4 and Methods), demonstrating the Onsager reciprocal relation between the AMPE and the AMSE.

Up to now, we have shown the AMPE-induced temperature modulation generated when  $H$  is along the  $x$  direction, where the Peltier coefficient for  $B_{L/R}$  ( $B_C$ ) is  $\Pi_{\perp}$  ( $\Pi_{\parallel}$ ) because of  $\mathbf{M} \perp \mathbf{J}_c$  ( $\mathbf{M} \parallel \mathbf{J}_c$ ). To confirm the symmetry of the AMPE, it is also important to demonstrate how the temperature modulation depends on the angle  $\theta$  between  $\mathbf{M}$  and  $\mathbf{J}_c$ . In Fig. 4, we show how the  $A_{\text{even}}$  and  $\phi_{\text{even}}$  images for the U-shaped Ni slab at  $J_c = 1.00$  A and  $|H| = 12.0$  kOe depend on  $\theta_{HH}$ , where  $\theta_{HH}$  is defined as the angle between  $H$  and the  $x$  direction. We found that the AMPE signal at  $\theta_{HH} = 90^\circ$ , where the Peltier coefficient for  $B_{L/R}$  ( $B_C$ ) is  $\Pi_{\parallel}$  ( $\Pi_{\perp}$ ) because of  $\mathbf{M} \parallel \mathbf{J}_c$  ( $\mathbf{M} \perp \mathbf{J}_c$ ), is opposite in sign to that at  $\theta_{HH} = 0^\circ$  (compare Fig. 4e, f with Fig. 4a, b), consistent with equation (2). We also checked that the temperature modulation at  $C_{L/R}$  disappears when  $\theta_{HH} \approx 45^\circ$  ( $\theta_{HH} = 40^\circ$  in the experiments shown in Fig. 4c, d), at which point both  $B_{L/R}$  and  $B_C$  have the same Peltier coefficient. Instead of temperature modulation at  $C_{L/R}$ , in this condition, a temperature gradient is generated along the  $x$  ( $y$ ) direction in  $B_{L/R}$  ( $B_C$ ). The generation of this temperature gradient can be explained by the transverse response due to  $\Pi_{\parallel} - \Pi_{\perp}$ , that is, a planar Ettingshausen effect (PEE) (see Methods). To our knowledge, the PEE has not previously been observed, but it is natural that the effect should appear concomitantly with the AMPE (recall that the transverse response due to the AMR induces the planar Hall effect<sup>29</sup> in ferromagnets). All the experimental results shown here are well reproduced by the numerical simulations, in which the charge-to-heat current conversion and the charge-current distribution in the U-shaped sample are taken into account (Extended Data Fig. 3 and Methods). The consistency between the experiments and simulations supports our interpretation that the current-induced temperature modulation with  $H$ -even dependence in the U-shaped Ni slab can be attributed to the AMPE and the PEE.

The AMPE observed here enables thermoelectric cooling/heating in a single material without junctions, providing new concepts in thermal management technologies for electronic and spintronic devices. Owing to the absence of junction structures, the thermoelectric properties of the AMPE can be redesigned simply by changing the shape of the ferromagnets or magnetization configurations. The unique characteristics of the AMPE potentially make it possible, for example, to construct compact

and  $|H| = 12.0$  kOe. To measure the  $H$ -angle dependence, the sample was rotated while the positions of the infrared camera and electromagnet were fixed. Therefore, the grey areas in **c** and **d** are out of the rectangular viewing area of the camera.

temperature controllers for nanoscale devices and integrated circuits, in which conventional Peltier junctions between different materials cannot be integrated because of their complicated structure and high production cost. The thermoelectric cooling/heating due to the AMPE can be produced not only in curved or bent ferromagnets but also in straight ferromagnets with non-uniform magnetization configurations, as shown in Fig. 1b. We have demonstrated this feature by applying a local magnetic field to a straight Ni slab (Extended Data Fig. 5 and Methods). The versatile thermoelectric capability of the AMPE is supported by the fact that the AMPE appears not only in Ni but also in other ferromagnetic materials (Extended Data Fig. 6 and Methods). However, as the present AMPE magnitude is insufficient to realize thermoelectric applications in commercial devices, further research will be needed to understand its microscopic mechanism, to explore materials with large anisotropy of Peltier coefficients and to engineer devices to maximize the AMPE.

Finally, we present a strategy for enhancing the AMPE. As the AMPE originates from the spin-orbit interaction acting on spin-polarized conduction electrons, doping of ferromagnetic materials with heavy elements that have strong spin-orbit interaction is useful. In fact, we found that  $\text{Ni}_{95}\text{Pt}_5$  and  $\text{Ni}_{95}\text{Pd}_5$  exhibit AMPE signals with magnitudes greater than that in pure Ni, where Pt and Pd possess strong spin-orbit interaction (Extended Data Fig. 6). The choice of host ferromagnetic materials is also important; although Ni and Ni-based alloys exhibit the AMPE, Fe does not show clear AMPE signals and  $\text{Ni}_{45}\text{Fe}_{55}$  shows only small signals (Extended Data Fig. 6), indicating that the parameters determining the Peltier coefficient of Fe and  $\text{Ni}_{45}\text{Fe}_{55}$  are hardly affected by the spin-orbit interaction. The strong material dependence of the AMPE indicates that the anisotropy of the Peltier coefficients could be enhanced further by designing appropriate combinations of host ferromagnets and doping materials. For high-performance and magnetic-field-free AMPE devices, it may also be interesting to investigate the AMPE in magnetic and spintronic materials with high magnetic anisotropy. Future research will also include the investigation of the potential roles of spin polarization ratio and magnon-electron drag<sup>13</sup> in the AMPE, because these factors may have a substantial impact on thermoelectric transport in ferromagnetic metals. We anticipate that the observation of the AMPE reported here will invigorate basic and applied research on electron transport phenomena and lead to considerable advances in thermoelectrics and spintronics.

## Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0143-x>.

Received: 3 November 2017; Accepted: 23 February 2018;  
Published online: 21 May 2018

1. Ashcroft, N. W. & Mermin, N. D. *Solid State Physics* (Saunders College, Philadelphia, 1976).
2. Kondepudi, D. & Prigogine, I. *Modern Thermodynamics: From Heat Engines to Dissipative Structures* (Wiley, Chichester, 1998).
3. Straube, H., Wagner, J.-M. & Breitenstein, O. Measurement of the Peltier coefficient of semiconductors by lock-in thermography. *Appl. Phys. Lett.* **95**, 052107 (2009).
4. Breitenstein, O., Warta, W. & Langenkamp, M. *Lock-in Thermography: Basics and Use for Evaluating Electronic Devices and Materials* 2nd edn (Springer, Berlin/Heidelberg, 2010).
5. Wid, O. et al. Investigation of the unidirectional spin heat conveyor effect in a 200 nm thin yttrium iron garnet film. *Sci. Rep.* **6**, 28233 (2016).
6. Daimon, S., Iguchi, R., Hioki, T., Saitoh, E. & Uchida, K. Thermal imaging of spin Peltier effect. *Nat. Commun.* **7**, 13754 (2016).
7. Wid, O. et al. Investigation of non-reciprocal magnon propagation using lock-in thermography. *J. Phys. D* **50**, 134001 (2017).
8. Uchida, K. et al. Enhancement of the spin Peltier effect in multilayers. *Phys. Rev. B* **95**, 184437 (2017).
9. Daimon, S., Uchida, K., Iguchi, R., Hioki, T. & Saitoh, E. Thermographic measurements of the spin Peltier effect in metal/yttrium-iron-garnet junction systems. *Phys. Rev. B* **96**, 024424 (2017).
10. Hirayama, Y., Iguchi, R., Miao, X.-F., Hono, K. & Uchida, K. High-throughput direct measurement of magnetocaloric effect based on lock-in thermography technique. *Appl. Phys. Lett.* **111**, 163901 (2017).
11. Bauer, G. E. W., Saitoh, E. & van Wees, B. J. Spin caloritronics. *Nat. Mater.* **11**, 391–399 (2012).
12. Boona, S. R., Myers, R. C. & Heremans, J. P. Spin caloritronics. *Energy Environ. Sci.* **7**, 885–910 (2014).
13. Watzman, S. J. et al. Magnon-drag thermopower and Nernst coefficient in Fe, Co, and Ni. *Phys. Rev. B* **94**, 144407 (2016).
14. Bozorth, R. M. Magnetoresistance and domain theory of iron–nickel alloys. *Phys. Rev.* **70**, 923–932 (1946).
15. Karplus, R. & Luttinger, J. M. Hall effect in ferromagnetics. *Phys. Rev.* **95**, 1154–1160 (1954).
16. McGuire, T. & Potter, R. Anisotropic magnetoresistance in ferromagnetic 3d alloys. *IEEE Trans. Magn.* **11**, 1018–1038 (1975).
17. Jan, J. P. in *Solid State Physics*, Vol. 5 (eds Seitz, F. & Turnbull, D.) 1–96 (1957).
18. Grannemann, G. N. & Berger, L. Magnon-drag Peltier effect in a Ni–Cu alloy. *Phys. Rev. B* **13**, 2072–2079 (1976).
19. Wegrowe, J.-E. et al. Anisotropic magnetothermopower: contribution of interband relaxation. *Phys. Rev. B* **73**, 134422 (2006).
20. Avery, A. D., Sultan, R., Basset, D., Wei, D. & Zink, B. L. Thermopower and resistivity in ferromagnetic thin films near room temperature. *Phys. Rev. B* **83**, 100401(R) (2011).
21. Mitdank, R. et al. Enhanced magneto-thermoelectric power factor of a 70 nm Ni-nanowire. *J. Appl. Phys.* **111**, 104320 (2012).
22. Avery, A. D., Pufall, M. R. & Zink, B. L. Observation of the planar Nernst effect in permalloy and nickel thin films with in-plane thermal gradients. *Phys. Rev. Lett.* **109**, 196602 (2012).
23. Avery, A. D., Pufall, M. R. & Zink, B. L. Determining the planar Nernst effect from magnetic-field-dependent thermopower and resistance in nickel and permalloy thin films. *Phys. Rev. B* **86**, 184408 (2012).
24. Schmid, M. et al. Transverse spin Seebeck effect versus anomalous and planar Nernst effects in permalloy thin films. *Phys. Rev. Lett.* **111**, 187201 (2013).
25. Reimer, O. et al. Quantitative separation of the anisotropic magnetothermopower and planar Nernst effect by the rotation of an in-plane thermal gradient. *Sci. Rep.* **7**, 40586 (2017).
26. Yamaguchi, A. et al. Real-space observation of current-driven domain wall motion in submicron magnetic wires. *Phys. Rev. Lett.* **92**, 077205 (2004).
27. Parkin, S. S. P., Hayashi, M. & Thomas, L. Magnetic domain-wall racetrack memory. *Science* **320**, 190–194 (2008).
28. Seki, T., Iguchi, R., Takanashi, K. & Uchida, K. Visualization of anomalous Ettingshausen effect in a ferromagnetic film: direct evidence of different symmetry from spin Peltier effect. *Appl. Phys. Lett.* **112**, 152403 (2018).
29. Ky, V. D. Planar Hall effect in ferromagnetic films. *Phys. Status Solidi* **26**, 565–569 (1968).

**Acknowledgements** We thank K. Masuda, Y. Miura, T. Seki, S. Takahashi, K. Sato and G. E. W. Bauer for discussions, and the Materials Processing Group, Materials Manufacturing and Engineering Station, National Institute for Materials Science, for technical support in sample preparation. This work was supported by CREST “Creation of Innovative Core Technologies for Nano-enabled Thermal Management” (JPMJCR1711), PRESTO “Phase Interfaces for Highly Efficient Energy Utilization” (JPMJPR12C1) and ERATO “Spin Quantum Rectification Project” (JPMJER1402) from JST, Japan; Grant-in-Aid for Scientific Research (A) (JP15H02012) and Grant-in-Aid for Scientific Research on Innovative Area “Nano Spin Conversion Science” (JP26103005) from JSPS KAKENHI, Japan; and the NEC Corporation. S.D. is supported by JSPS through a research fellowship for young scientists (JP16J02422).

**Reviewer information** Nature thanks S. Boona and A. Fert for their contribution to the peer review of this work.

**Author contributions** K.U. planned and supervised the study, designed the experiments, prepared the samples, collected and analysed the data, and prepared the manuscript. S.D. and R.I. performed the numerical calculations. K.U., S.D., R.I. and E.S. discussed the results, developed the explanation of the experiments and commented on the manuscript.

**Competing interests** The authors declare no competing interests.

#### Additional information

**Extended data** is available for this paper at <https://doi.org/10.1038/s41586-018-0143-x>.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>.

**Correspondence and requests for materials** should be addressed to K.-i.U.  
**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## METHODS

**Sample preparation for the measurements of the anisotropic magneto-Peltier effect.** The polycrystalline Ni, Ni<sub>45</sub>Fe<sub>55</sub> and Fe slabs used in this study are commercially available from the Nilaco Corporation, Japan. The polycrystalline Ni<sub>95</sub>Pt<sub>5</sub> and Ni<sub>95</sub>Pd<sub>5</sub> slabs used for measuring the data in Extended Data Fig. 6 were prepared by means of a melting method by Kojundo Chemical Laboratory Co. Ltd, Japan. For measuring the AMPE, except for the experiments shown in Extended Data Fig. 5, the ferromagnetic metal slabs were cut into a U-shaped structure by wire electrical discharge machining. The length along the  $y$  direction, total width along the  $x$  direction, width of the areas B<sub>L/C/R</sub> and thickness of the U-shaped Ni, Ni<sub>45</sub>Fe<sub>55</sub> and Fe (Ni<sub>95</sub>Pt<sub>5</sub> and Ni<sub>95</sub>Pd<sub>5</sub>) slabs are 10.0 (12.0), 1.0 (1.0), 0.2 (0.2) and 0.5 (0.5) mm, respectively. During the LIT measurements, the sample was fixed on a plastic plate with low thermal conductivity ( $<0.3 \text{ W m}^{-1} \text{ K}^{-1}$ ) to reduce the heat loss due to thermal conduction as much as possible.

**Calibration of thermal images.** The infrared radiation intensity  $I$  detected by the infrared camera is converted into temperature  $T$  information by measuring the  $T$  dependence of  $I$  for a black-ink-coated Ni plate at thermal equilibrium. Here, the  $I$ -to- $T$  conversion for LIT images is determined by the differential relation  $\Delta T_{\text{if}} = dT/dI|_T \Delta I_{\text{if}}$ , where  $\Delta T_{\text{if}}$  and  $\Delta I_{\text{if}}$  denote the lock-in responses of the temperature and infrared radiation intensity, respectively.

**Numerical calculations.** To confirm the symmetry of the AMPE and the PEE and to quantitatively estimate the anisotropy of the Peltier coefficient, we carried out numerical calculations with a three-dimensional finite element method<sup>30</sup>. The model system used for the calculations is a U-shaped ferromagnetic metal, the size and shape of which are based on the U-shaped Ni slab used in the experiments. By performing multiphysics simulations, we take the distribution of a charge current applied to the ferromagnetic metal and of the resultant heat source due to the charge-to-heat current conversion into account in the calculations of the temperature modulation.

In an isotropic ferromagnet, when its magnetization is in the  $x$ - $y$  plane, the matrix representation of the AMPE and the PEE can be given by

$$\begin{pmatrix} j_{q,x}^{\text{AMPE}} \\ j_{q,y}^{\text{AMPE}} \\ j_{q,z}^{\text{AMPE}} \end{pmatrix} = \left[ \Pi_{\perp} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + (\Pi_{\parallel} - \Pi_{\perp}) \begin{pmatrix} \cos^2 \theta & \cos \theta \sin \theta & 0 \\ \cos \theta \sin \theta & \sin^2 \theta & 0 \\ 0 & 0 & 0 \end{pmatrix} \right] \begin{pmatrix} j_{c,x} \\ j_{c,y} \\ j_{c,z} \end{pmatrix} \quad (3)$$

where  $j_{q,i}^{\text{AMPE}}$  represents the heat-current density due to the AMPE (diagonal components) and the PEE (off-diagonal components) along the  $i$  ( $=x, y, z$ ) axis,  $j_{c,i}$  the  $i$  component of the charge-current-density vector  $\mathbf{j}_c$ , and  $\theta$  the angle between the directions of the charge current  $\mathbf{j}_{c,x}$  and the magnetization in the ferromagnetic metal (see also equation (2)). Here,  $\mathbf{j}_c = -\sigma \nabla \Phi$  is determined by solving Laplace's equation:  $\nabla^2 \Phi = 0$ , where  $\sigma$  is the electrical conductivity and  $\Phi$  is the electrostatic potential. In the calculations, the difference in the  $\Phi$  values at the ends of the two legs of the U-shaped structure is calibrated to reproduce the magnitude of the charge current applied in the experiments. In the calculations, we neglected the small contributions from the AMR<sup>16</sup> and the Maggi-Righi-Leduc effect<sup>12</sup> in the ferromagnetic metal. The capacitive contribution is also not taken into account as the frequency of the applied current is low ( $\leq 25.0$  Hz in the experiments).

The amplitude of the temperature modulation  $A$  and its phase delay  $\phi$  are determined by solving the heat diffusion equation in the frequency domain. By substituting the temperature  $T(t)$  with  $\int A e^{i(2\pi f t - \phi)} df$  in the heat diffusion equation, we obtained  $2\pi i f C D A e^{-i\phi} = \nabla \cdot (\kappa \nabla A e^{-i\phi}) + \dot{Q}_j$  in the frequency domain, where  $C$  is the specific heat,  $D$  the density,  $\kappa$  the thermal conductivity of the ferromagnetic metal, and  $\dot{Q}_j = -\nabla \cdot \mathbf{j}_q^{\text{AMPE}}$  the amplitude of a source term induced by the AMPE due to the alternating charge current  $\tilde{j}_c e^{2\pi i f t}$  (note that a tilde is attached to physical quantities in the frequency domain). As the boundary condition,  $A = 0$  is assumed at the ends of the two legs of the U-shaped structure. At the other surface boundaries, we set the following condition:  $(-\kappa \nabla A e^{-i\phi} + \tilde{j}_q^{\text{AMPE}}) \cdot \mathbf{n} = \tilde{j}_q^{\text{air}}$  with  $\mathbf{n}$  and  $\tilde{j}_q^{\text{air}} = \alpha_{\text{air}} A e^{-i\phi}$  being, respectively, the surface-normal vector (directed outwards) and the heat loss to air. For the simulations, we assumed  $C = 443 \text{ J kg}^{-1} \text{ K}^{-1}$ ,  $D = 8,908 \text{ kg m}^{-3}$ ,  $\kappa = 90.7 \text{ W m}^{-1} \text{ K}^{-1}$ , which are reference values for Ni (refs <sup>31,32</sup>), and  $\alpha_{\text{air}} = 10 \text{ W m}^{-2} \text{ K}^{-1}$  (ref. <sup>6</sup>).

We estimated the anisotropy of the Peltier coefficient,  $\Pi_{\parallel} - \Pi_{\perp}$ , for our Ni sample by comparing the calculated temperature modulation with the experimental results. In Extended Data Fig. 3c, d, we show the calculated  $A$  and  $\phi$  distributions on the surface of the ferromagnetic metal for various  $f$  values at  $\Pi_{\parallel} - \Pi_{\perp} = 0.11 \times 10^{-3} \text{ V}$  under the rectangular alternating charge current of  $J_c = 1.00 \text{ A}$  (note that, in the calculations, the current value of  $(4/\pi)J_c$  is used as this is the amplitude of the first harmonic trigonometric component in the rectangularly modulated alternating current with rectangular amplitude  $J_c$ ). We found that a clear temperature modulation is generated around the corners of the U-shaped ferromagnetic

metal. With decreasing  $f$ , the temperature modulation broadens because of thermal diffusion, and its magnitude increases. The  $f$  dependence of  $A$  and  $\phi$ , taken from the areas C<sub>L</sub> and C<sub>R</sub> on the surface of the ferromagnetic metal, is consistent with the experimental results (Extended Data Fig. 3a, b).

To understand the symmetry of the AMPE and the PEE, we next calculated the  $\theta$  dependence of the  $A$  and  $\phi$  distributions at  $f = 25.0$  Hz. As shown in Extended Data Fig. 3e, f, the spatial distribution of the temperature modulation governed by equation (3) depends strongly on the magnetization direction and well reproduces the  $H$ -angle dependence of the LIT images in Fig. 4.

**Measurements of the anisotropic magneto-Seebeck effect.** In Extended Data Fig. 4, we report the observation of the AMSE in a straight Ni slab with a rectangular shape. The lengths of the Ni slab along the  $x$ ,  $y$  and  $z$  directions are 0.5, 10.0 and 1.0 mm, respectively. To apply a temperature gradient along the  $y$  direction, one end of the Ni slab was attached to a chip heater and the other end to a heat bath. The temperature difference  $\Delta T$  between the ends of the Ni slab (coated with insulating black ink) was measured with the infrared camera. We measured the electric voltage between the ends of the Ni slab at finite  $\Delta T$  values while applying  $\mathbf{H}$  in the direction perpendicular to or parallel to the temperature gradient  $\nabla T$  at room temperature (see schematic illustrations in Extended Data Fig. 4a).

Extended Data Fig. 4b, c shows the voltage  $\Delta V$  between the ends of the Ni slab as a function of  $H$  for various values of  $\Delta T$ , where  $\Delta V$  was obtained by removing the offset due to  $H$ -independent thermopower. The  $\Delta V$  signals exhibit clear  $H$ -even dependence; when  $\mathbf{H} \perp \nabla T$  ( $\mathbf{H} \parallel \nabla T$ ), the magnitude of  $\Delta V$  increases (decreases) with increasing  $|H|$  and saturates around the saturation field of the Ni slab, consistent with the feature of the AMSE<sup>12</sup>. The field-induced change in  $\Delta V$  is proportional to  $\Delta T$  (Extended Data Fig. 4a), and its thermopower was observed to be  $\Delta S_{\perp(\parallel)} = -\Delta V / \Delta T = -0.2 \times 10^{-6} \text{ V K}^{-1}$  ( $0.4 \times 10^{-6} \text{ V K}^{-1}$ ) at  $|H| = 6.0 \text{ kOe}$  (1.0 kOe) for the  $\mathbf{H} \perp \nabla T$  ( $\mathbf{H} \parallel \nabla T$ ) configuration. Therefore, the anisotropy of the Seebeck coefficient for our Ni slab is estimated to be  $[(\Delta S_{\parallel} - \Delta S_{\perp})/S] = 3\%$  of the conventional Seebeck coefficient,  $S = -20 \times 10^{-6} \text{ V K}^{-1}$  (ref. <sup>21</sup>), at room temperature. The observed magnitude and sign of the AMSE are consistent with those of the AMPE, demonstrating the Onsager reciprocal relation between these phenomena (note that the small difference in the magnitude of the anisotropy between the AMSE and the AMPE might be attributed to the heat loss from the Ni slab to the plastic plate in the AMPE measurements).

In Extended Data Fig. 6c, we show the  $H$  dependence of  $\Delta V / \Delta T$  not only for Ni but also for various ferromagnetic metals, Ni<sub>95</sub>Pt<sub>5</sub>, Ni<sub>95</sub>Pd<sub>5</sub>, Ni<sub>45</sub>Fe<sub>55</sub> and Fe, measured by the aforementioned procedure. We found that the magnitude of the AMSE signals in Ni<sub>95</sub>Pt<sub>5</sub> and Ni<sub>95</sub>Pd<sub>5</sub> (Ni<sub>45</sub>Fe<sub>55</sub>) is greater (smaller) than that in Ni, and the sign of the signals in these materials is the same as that in Ni. The AMSE signal in Fe was observed to be negligibly small. These behaviours are consistent with the experimental results of the AMPE shown in Extended Data Fig. 6a, b.

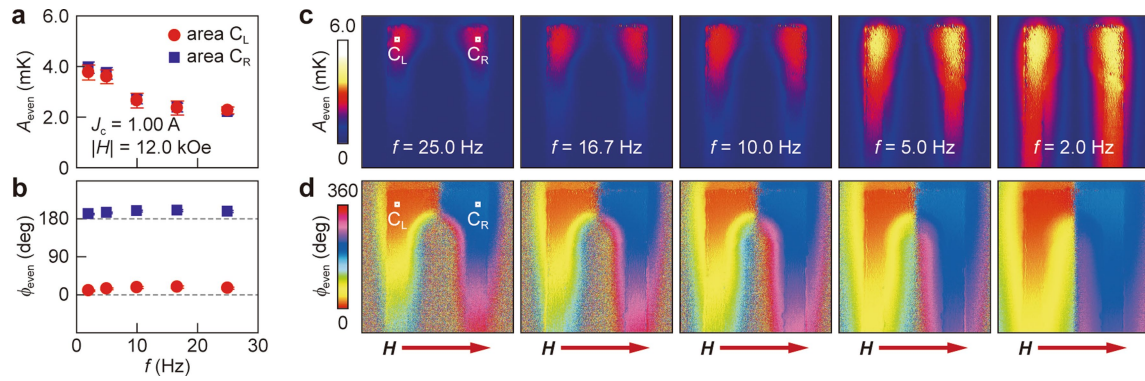
**Anisotropic magneto-Peltier effect induced by local magnetic fields.** The AMPE signals can be generated even in the absence of curved or bent structures by applying local magnetic fields to a ferromagnet. Extended Data Fig. 5a shows a schematic illustration of the experimental configuration for measuring the local-field-induced AMPE. We applied the local magnetic field along the  $x$  direction to the centre part of a straight Ni slab by using an electromagnet with narrow magnetic pole pieces, where the lengths of the slab along the  $x$ ,  $y$  and  $z$  directions are 0.2, 20.0 and 1.0 mm, respectively, and the gap and width of the pole pieces are 1.0 mm. When only the centre part of the Ni slab is magnetized and  $J_c$  is applied along the  $y$  direction, the AMPE signal should appear in the areas adjacent to the magnetized region because the magnetized region possesses  $\Pi_{\parallel}$ , but the remaining non-magnetized regions possess the  $\theta$ -averaged  $\Pi$ . Judging from the results shown in the main text, the  $\theta$ -averaged  $\Pi$  is greater than  $\Pi_{\perp}$ , inducing heat absorption and release as schematically depicted in Extended Data Fig. 5a. To detect the AMPE induced by the local magnetic field, we measured the temperature distribution on the  $0.2 \times 20.0 \text{ mm}^2$  surface of the straight Ni slab by means of the LIT method, applying a rectangularly modulated alternating charge current along the  $y$  direction with  $J_c = 1.00 \text{ A}$ ,  $f = 25.0$  Hz and zero d.c. offset. The magnitude of the local magnetic field at the centre of the sample is fixed at  $|H| = 6.0 \text{ kOe}$ , comparable to the saturation field for the straight Ni slab along the  $x$  direction, so that the magnetization aligns along the  $\mathbf{H}$  direction only in the area between the pole pieces.

To check the local field distribution, we first show the  $A_{\text{odd}}$  and  $\phi_{\text{odd}}$  images for the straight Ni slab. As shown in Extended Data Fig. 5b, c, a clear AEE signal is generated around the centre of the Ni slab, and its magnitude decreases with distance from the pole pieces, indicating that the magnetization of the Ni slab aligns along the  $x$  direction only near the centre. This non-uniform magnetization configuration is suitable for exploiting the AMPE.

Extended Data Fig. 5d, e shows the  $A_{\text{even}}$  and  $\phi_{\text{even}}$  images for the straight Ni slab. The temperature modulation with  $H$ -even dependence was observed to exhibit two peaks and appear only in the intermediate areas between the magnetized and non-magnetized regions (Extended Data Fig. 5d). The sign of the temperature modulation shown in Extended Data Fig. 5e indicates that heat is absorbed (released)

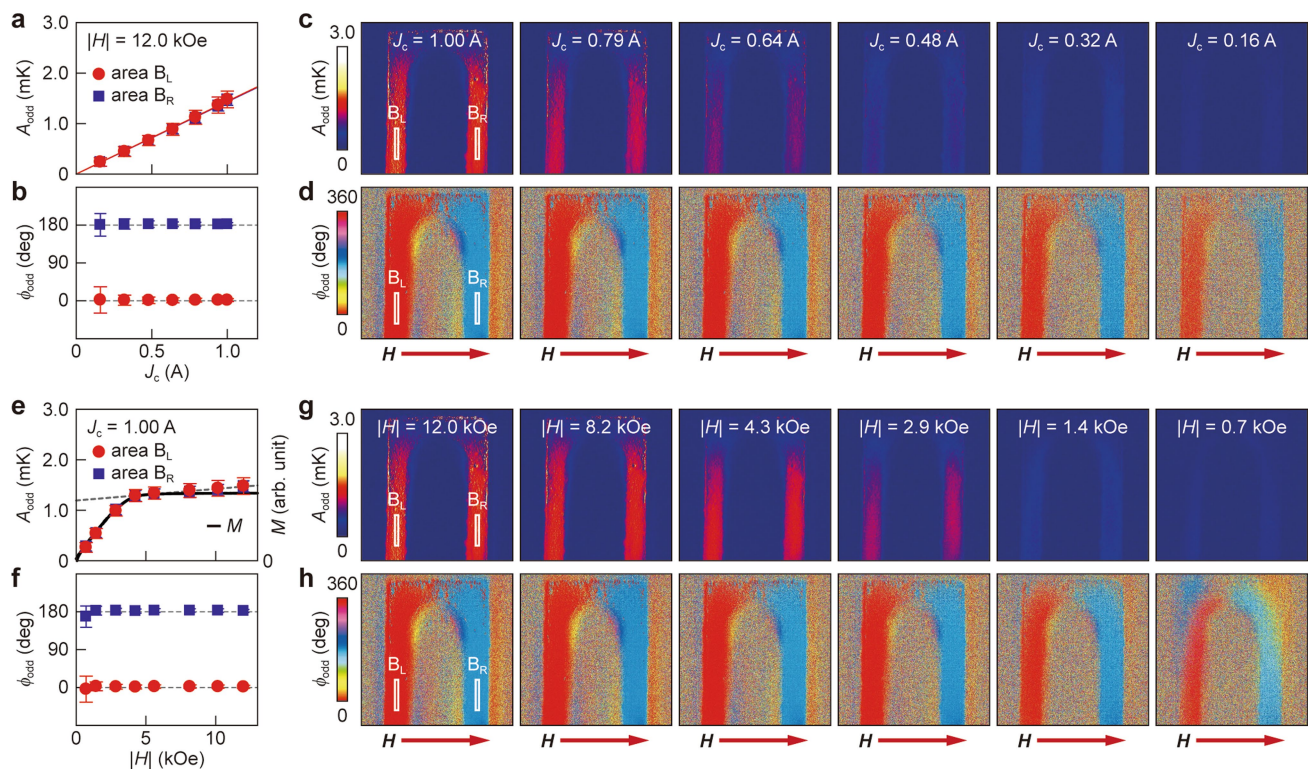
when the charge current flows from the magnetized (non-magnetized) region to the non-magnetized (magnetized) region, which is consistent with the behaviour expected for the AMPE induced by the local magnetic field (Extended Data Fig. 5a). **Data availability.** The data that support the findings of this study are available from the corresponding author upon reasonable request.

30. Hecht, F. New development in freefem++. *J. Numer. Math.* **20**, 251–265 (2012).
31. Desai, P. D. Thermodynamic properties of nickel. *Int. J. Thermophys.* **8**, 763–780 (1987).
32. Ho, C. Y., Powell, R. W. & Liley, P. E. *Thermal Conductivity of the Elements: A Comprehensive Review* (AIP/ACS, New York, 1974).



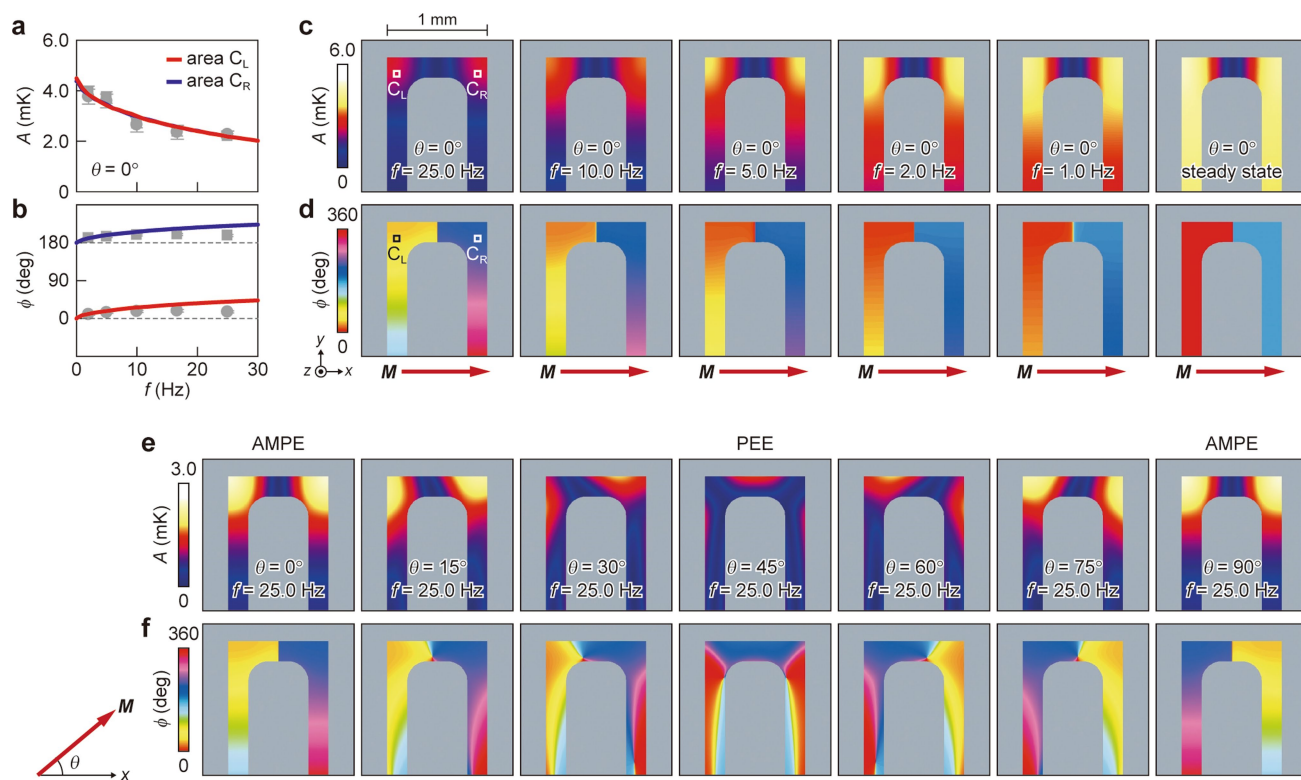
**Extended Data Fig. 1 | Lock-in frequency dependence of anisotropic magneto-Peltier effect. a, b,** Frequency dependence of  $A_{\text{even}}$  and  $\phi_{\text{even}}$  on the areas  $C_{L/R}$  of the U-shaped Ni slab at  $J_c = 1.00$  A and  $|H| = 12.0$  kOe. **c, d,**  $A_{\text{even}}$  and  $\phi_{\text{even}}$  images for the U-shaped Ni slab at  $J_c = 1.00$  A and  $|H| = 12.0$  kOe for various values of  $f$ . The data points in **a** and **b** are respectively obtained by averaging the  $A_{\text{even}}$  and  $\phi_{\text{even}}$  values over the areas defined by the squares with size  $11 \times 11$  pixels in the leftmost images of **c** and **d**. The error bars represent the standard deviation of the data in the corresponding squares. In Figs. 2–4, to determine the positions of heat sources induced by the AMPE, the lock-in frequency was fixed at the maximum value,  $f = 25.0$  Hz, because the temperature broadening due to thermal diffusion is reduced by increasing  $f$ . However, in general, the

temperature distribution obtained from the LIT images at high  $f$  values is different from the steady-state distribution. To discuss the magnitude of the AMPE, it is important to observe the temperature modulation at nearly steady-state conditions. As shown in **c** and **d**, the AMPE signal around the corners  $C_{L/R}$  of the U-shaped structure broadens with decreasing  $f$  owing to thermal diffusion. Although the magnitude of the AMPE signal at  $C_{L/R}$  is  $A_{\text{even}} = 2.2$  mK at  $f = 25.0$  Hz, it increases to  $A_{\text{even}} = 3.9$  mK at  $f = 2.0$  Hz, which is closer to the value at the steady state. In Extended Data Fig. 3, these experimental results are compared with the results of numerical calculations to estimate the anisotropy of the Peltier coefficient for Ni quantitatively.



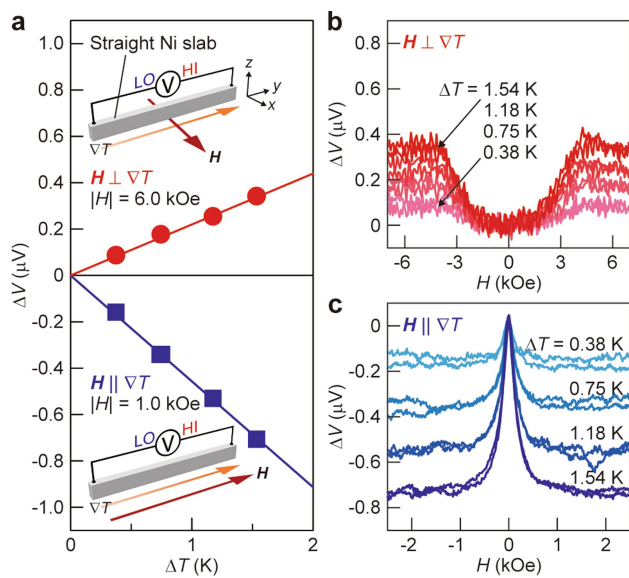
**Extended Data Fig. 2 | Charge current and magnetic field dependences of anomalous Ettingshausen effect.** **a, b,**  $J_c$  dependence of  $A_{\text{odd}}$  and  $\phi_{\text{odd}}$  on the areas  $B_{L/R}$  of the U-shaped Ni slab at  $|H| = 12.0$  kOe (red circles and blue squares). The solid lines in **a** represent the linear fits to the data. **c, d,**  $A_{\text{odd}}$  and  $\phi_{\text{odd}}$  images for the U-shaped Ni slab at  $|H| = 12.0$  kOe for various values of  $J_c$ . **e, f,**  $|H|$  dependence of  $A_{\text{odd}}$  and  $\phi_{\text{odd}}$  on  $B_{L/R}$  at  $J_c = 1.00$  A (red circles and blue squares) and the magnetization  $M$  curve of the Ni slab (black line). **g, h,**  $A_{\text{odd}}$  and  $\phi_{\text{odd}}$  images at  $J_c = 1.00$  A for various values of  $|H|$ . The data points in **a** and **b** (**e** and **f**) are respectively obtained by averaging the  $A_{\text{odd}}$  and  $\phi_{\text{odd}}$  values on the areas defined by the rectangles with size  $101 \times 11$  pixels in the leftmost images of **c** and **d** (**g** and **h**). The error bars represent the standard deviation of the data in

the corresponding rectangles. In the experimental configuration shown in Fig. 2b in addition to the AMPE, the heat current is generated along the  $z$  direction due to the AEE in  $B_{L/R}$  of the U-shaped Ni slab, where  $\mathbf{M}$  (along the  $x$  direction) is perpendicular to  $\mathbf{J}_c$  (along the  $y$  direction). The  $A_{\text{odd}}$  and  $\phi_{\text{odd}}$  images shown here clearly reflect the symmetry of the AEE. The temperature modulation with the  $H$ -odd dependence appears only on  $B_{L/R}$  and its sign is reversed between  $B_L$  and  $B_R$ , where the  $\mathbf{J}_c$  direction in  $B_L$  is opposite to that in  $B_R$ . The magnitude of the temperature modulation with the  $H$ -odd dependence is proportional to the charge current and varies in response to the magnetization process of the Ni slab. Here, we observed not only the AEE but also the small  $H$ -linear contribution coming from the ordinary Ettingshausen effect, as shown in **e** (grey dotted line).



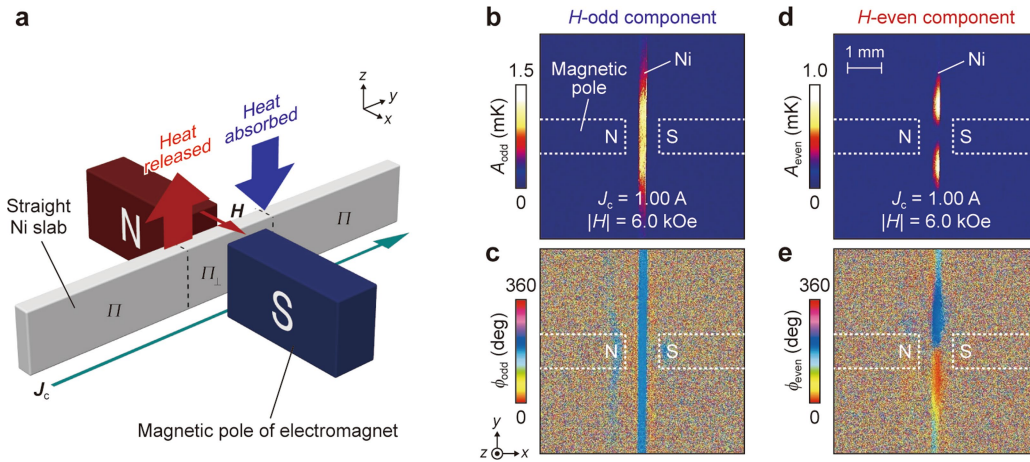
**Extended Data Fig. 3 | Numerical calculations of anisotropic magneto-Peltier effect.** **a, b**, Calculated  $f$  dependence of  $A$  and  $\phi$  on the areas  $C_{L/R}$  of the U-shaped ferromagnetic metal model at  $\theta = 0^\circ$  (red and blue lines). The grey plots are the experimental results shown in Extended Data Fig. 1a, b. **c, d**, Calculated  $A$  and  $\phi$  images for the U-shaped ferromagnetic metal model at  $\theta = 0^\circ$  for various values of  $f$ . The data points in **a** and

**b** are respectively obtained by averaging the  $A$  and  $\phi$  values over the areas defined by the squares in the leftmost images of **c** and **d**. **e, f**, Calculated  $A$  and  $\phi$  images at  $f = 25.0$  Hz for various values of  $\theta$ . The calculated temperature distributions reproduce well the observed  $H$ -angle dependence of the LIT images shown in Fig. 4.



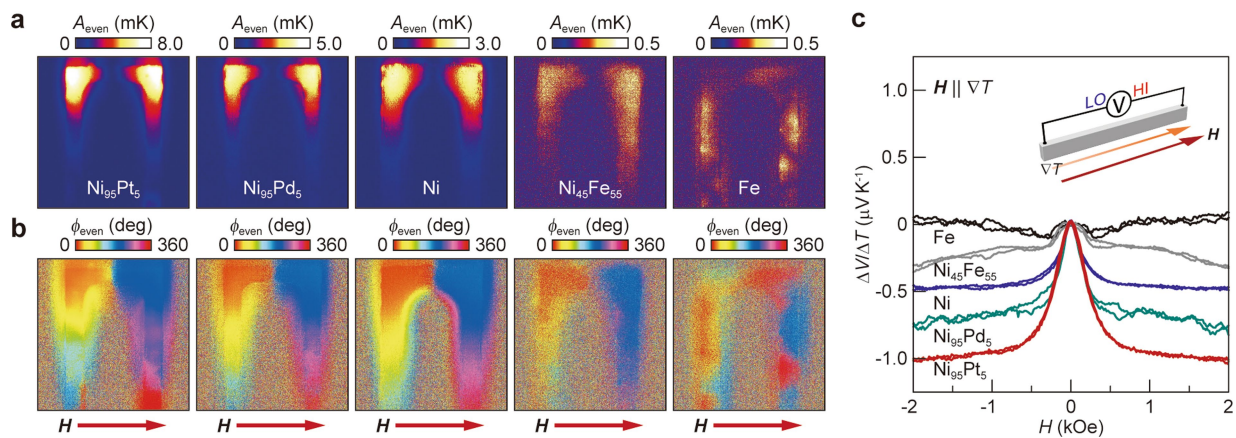
**Extended Data Fig. 4 | Anisotropic magneto-Seebeck effect in Ni.**

**a**, Temperature-difference  $\Delta T$  dependence of the voltage  $\Delta V$  between the ends of the straight Ni slab at  $|H| = 6.0$  kOe (1.0 kOe), measured when  $H$  was applied in the direction perpendicular to (parallel to) the temperature gradient  $\nabla T$ . In the  $\Delta V$  data, the offset due to  $H$ -independent thermopower is subtracted. **b**, **c**,  $H$  dependence of  $\Delta V$  in the straight Ni slab for various values of  $\Delta T$ , measured for  $H \perp \nabla T$  (**b**) and  $H \parallel \nabla T$  (**c**). The difference in the shape of the  $H$ - $\Delta V$  curves between the data in **b** and **c** is attributed to the shape magnetic anisotropy in the Ni slab. We confirmed that the magnitude of the AMSE signal in the  $H \parallel \nabla T$  configuration is twice as large as that in the  $H \perp \nabla T$  configuration, in a similar manner to the AMR<sup>16</sup>.



**Extended Data Fig. 5 | Anisotropic magneto-Peltier effect induced by local magnetic fields.** **a**, Experimental configuration for measuring the AMPE induced by the local magnetic field  $H$ . The field was applied near the centre of the straight Ni slab.  $\Pi$  denotes the Peltier coefficient of

non-magnetized Ni, that is, the  $\theta$ -averaged Peltier coefficient. **b**, **c**,  $A_{\text{odd}}$  and  $\phi_{\text{odd}}$  images for the straight Ni slab at  $J_c = 1.00$  A and  $|H| = 6.0$  kOe. **d**, **e**,  $A_{\text{even}}$  and  $\phi_{\text{even}}$  images at  $J_c = 1.00$  A and  $|H| = 6.0$  kOe.



**Extended Data Fig. 6 | Anisotropic magneto-Peltier and Seebeck effects in various ferromagnetic metals. a, b,**  $A_{\text{even}}$  and  $\phi_{\text{even}}$  images for the U-shaped  $\text{Ni}_{95}\text{Pt}_5$ ,  $\text{Ni}_{95}\text{Pd}_5$ ,  $\text{Ni}$ ,  $\text{Ni}_{45}\text{Fe}_{55}$  and  $\text{Fe}$  slabs at  $J_c = 1.00$  A,  $|H| = 12.0$  kOe, and  $f = 25.0$  Hz. The  $\text{Ni}_{95}\text{Pt}_5$  and  $\text{Ni}_{95}\text{Pd}_5$  slabs were found to exhibit clear AMPE signals with greater magnitude than that in  $\text{Ni}$ . Although the  $\text{Ni}_{45}\text{Fe}_{55}$  slab also exhibits the AMPE, its magnitude is smaller than that in  $\text{Ni}_{95}\text{Pt}_5$ ,  $\text{Ni}_{95}\text{Pd}_5$  and  $\text{Ni}$ . The  $\text{Fe}$  slab does not show clear AMPE signals; the patchy patterns in the  $A_{\text{even}}$  and  $\phi_{\text{even}}$  images for  $\text{Fe}$  may arise because the magnetization of the U-shaped  $\text{Fe}$  slab is not

saturated at  $|H| = 12.0$  kOe, the maximum magnetic field available with our electromagnet (note that, in the AMPE measurements,  $H$  was applied along the hard axis due to the shape magnetic anisotropy). **c,**  $H$  dependence of  $\Delta V / \Delta T$  in the straight  $\text{Ni}_{95}\text{Pt}_5$ ,  $\text{Ni}_{95}\text{Pd}_5$ ,  $\text{Ni}$ ,  $\text{Ni}_{45}\text{Fe}_{55}$  and  $\text{Fe}$  slabs, measured when  $H \parallel \nabla T$ . As is the case for the AMPE, the  $\text{Ni}_{95}\text{Pt}_5$  and  $\text{Ni}_{95}\text{Pd}_5$  slabs exhibit clear AMSE signals with greater magnitude than that in  $\text{Ni}$ . In these AMSE measurements, the magnetization of the ferromagnetic metal slabs easily aligns along the  $H$  direction because  $H$  was applied along the longest direction of the slabs.

# Potential enthalpic energy of water in oils exploited to control supramolecular structure

Nathan J. Van Zee<sup>1</sup>, Beatrice Adelizzi<sup>1</sup>, Mathijs F. J. Mabesoone<sup>1</sup>, Xiao Meng<sup>1</sup>, Antonio Aloï<sup>1,2</sup>, R. Helen Zha<sup>1</sup>, Martin Lutz<sup>3</sup>, Ivo A. W. Filot<sup>1,4</sup>, Anja R. A. Palmans<sup>1</sup> & E. W. Meijer<sup>1\*</sup>

**Water directs the self-assembly of both natural<sup>1,2</sup> and synthetic<sup>3–9</sup> molecules to form precise yet dynamic structures. Nevertheless, our molecular understanding of the role of water in such systems is incomplete, which represents a fundamental constraint in the development of supramolecular materials for use in biomaterials, nanoelectronics and catalysis<sup>10</sup>. In particular, despite the widespread use of alkanes as solvents in supramolecular chemistry<sup>11,12</sup>, the role of water in the formation of aggregates in oils is not clear, probably because water is only sparingly miscible in these solvents—typical alkanes contain less than 0.01 per cent water by weight at room temperature<sup>13</sup>. A notable and unused feature of this water is that it is essentially monomeric<sup>14</sup>. It has been determined previously<sup>15</sup> that the free energy cost of forming a cavity in alkanes that is large enough for a water molecule is only just compensated by its interaction with the interior of the cavity; this cost is therefore too high to accommodate clusters of water. As such, water molecules in alkanes possess potential enthalpic energy in the form of unrealized hydrogen bonds. Here we report that this energy is a thermodynamic driving force for water molecules to interact with co-dissolved hydrogen-bond-based aggregates in oils. By using a combination of spectroscopic, calorimetric, light-scattering and theoretical techniques, we demonstrate that this interaction can be exploited to modulate the structure of one-dimensional supramolecular polymers.**

We discovered the importance of this principle in exploring the self-assembly of enantiopure **1** (Fig. 1a). In the bulk, **1** forms helical liquid crystalline structures at 20 °C with dimensions typical of one-dimensional aggregates<sup>16</sup> (Extended Data Fig. 1). As seen from super-resolution fluorescence and atomic force microscopy (AFM) images (Fig. 1b, c, respectively, and Extended Data Fig. 2), **1** also forms one-dimensional aggregates when diluted in oils such as methylcyclohexane (MCH). At micromolar concentrations in MCH, aggregates of **1** exhibit a preferred helicity, as observed by variable temperature circular dichroism (VT-CD) spectroscopy. However, surprisingly, the sign of the Cotton effect depends on the temperature of the solution. Early experiments were marred by seemingly inexplicable irreproducibility of the critical temperatures at which the transitions between these different helicities occur. We discovered the underlying cause by unexpectedly drying a solution of **1** in the nitrogen-purged atmosphere of the sample holder of the CD spectrometer (see Extended Data Fig. 3 for an analogous experiment). This dried sample no longer exhibited the distinctive helical transitions.

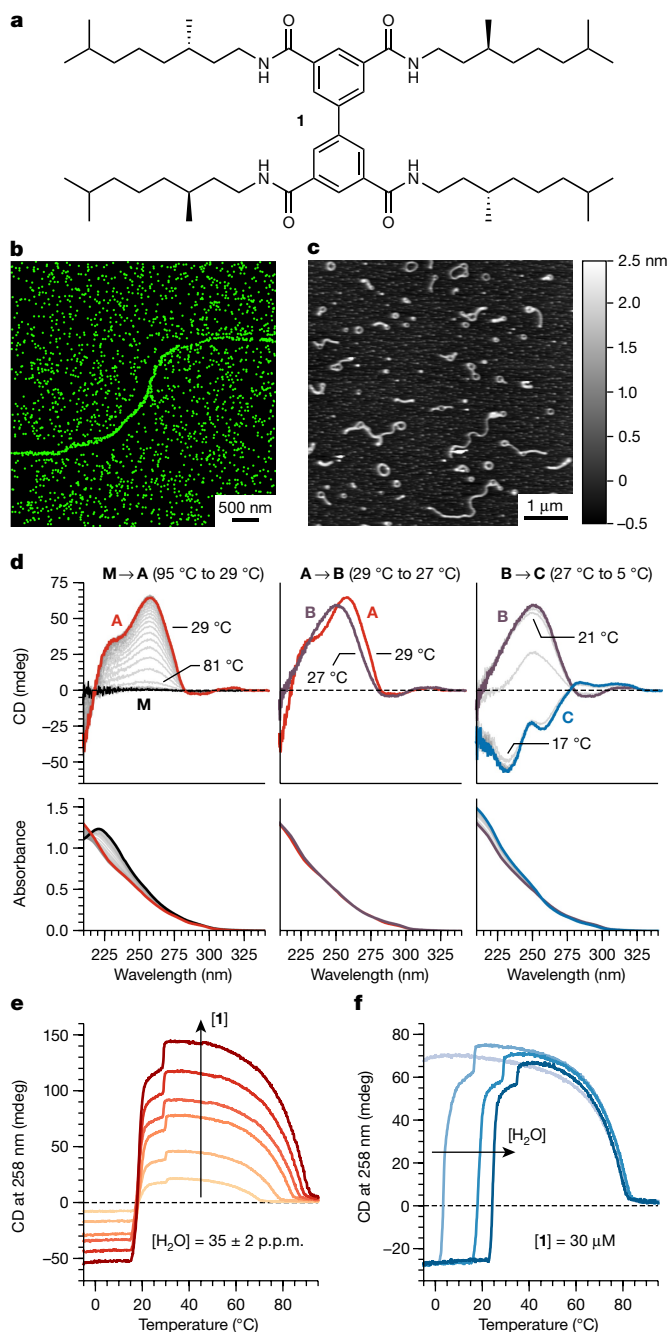
It became clear that unintended fluctuations in water content accounted for these initial inconsistencies, which motivated us to investigate the effect of water in supramolecular aggregation in oils in general, and its role in determining the helicity of aggregates of **1** in particular. We assessed the self-assembly of **1** in MCH ([**1**] = 30 μM) with 35 p.p.m. water (measured by Karl Fischer titration) by VT-CD and ultraviolet spectroscopies between 95 °C and 5 °C while cooling at 60 °C h<sup>–1</sup> (Fig. 1d). At temperatures greater than 81 °C, **1** exhibits an ultraviolet absorption maximum at 221 nm and no CD signal,

which indicates that it is molecularly dissolved. Upon cooling to 81 °C, **1** self-assembles to form chiral aggregates, **A**, which have a positive bisignate Cotton effect at 258 nm along with a blue-shifted ultraviolet spectrum. Cooling from 81 °C to 29 °C results in a rapid increase in intensity of the CD signal, which suggests that **A** forms through a cooperative supramolecular polymerization of **1**. An unprecedentedly abrupt transition is then observed when the solution is cooled from 29 °C to 27 °C, resulting in the conversion of **A** into a second chiral aggregate, **B**, with a positive Cotton effect at 250 nm and essentially the same ultraviolet spectrum as that of **A**. Continued cooling to 21 °C results in a slight decrease in the intensity of the CD signal of **B**, but cooling just below 21 °C gives rise to the rapid transformation of **B** into a third chiral aggregate, **C**, which exhibits a negative Cotton effect at 238 nm; further cooling does not initiate additional helical transitions.

Using VT-CD spectroscopy, we next investigated how the concentrations of both **1** and water affect the critical temperatures of these transitions. The progress of self-assembly was followed by monitoring the intensity of the Cotton effect at 258 nm as solutions were cooled from 95 °C to –5 °C at 60 °C h<sup>–1</sup>. The formation of **A** is dependent on the concentration of **1**, as its critical elongation temperature decreases from 91 °C to 70 °C as the concentration of **1** decreases from 59 μM to 10 μM (Fig. 1e). A van 't Hoff analysis reveals that the enthalpy of elongation is –86 kJ (mol **1**)<sup>–1</sup> (Extended Data Fig. 4a), consistent with this process being driven by the formation of four hydrogen bonds per molecule of **1**<sup>17,18</sup>. The transformations of **A** → **B** and **B** → **C** are not sensitive to the concentration of **1** and take place over a narrow temperature range. Moreover, these transitions exhibit a marked dependence on the concentration of co-dissolved water (Fig. 1f). At 47 p.p.m. water, **A** → **B** and **B** → **C** occur at 35 °C and 25 °C, respectively. As the water content decreases, they occur at progressively lower temperatures, which suggests that water binds to the structures of **B** and **C**. Van 't Hoff plots reveal that the enthalpies of **A** → **B** and **B** → **C** are about –21 and –26 kJ (mol H<sub>2</sub>O)<sup>–1</sup>, respectively, which indicates that the enthalpic driving force for water to bind is essentially the same in each transition (Extended Data Fig. 4b, c). At 8 p.p.m. water, only **A** is formed even after cooling to –5 °C, but it should be noted that water is still present in over tenfold molar excess relative to **1**.

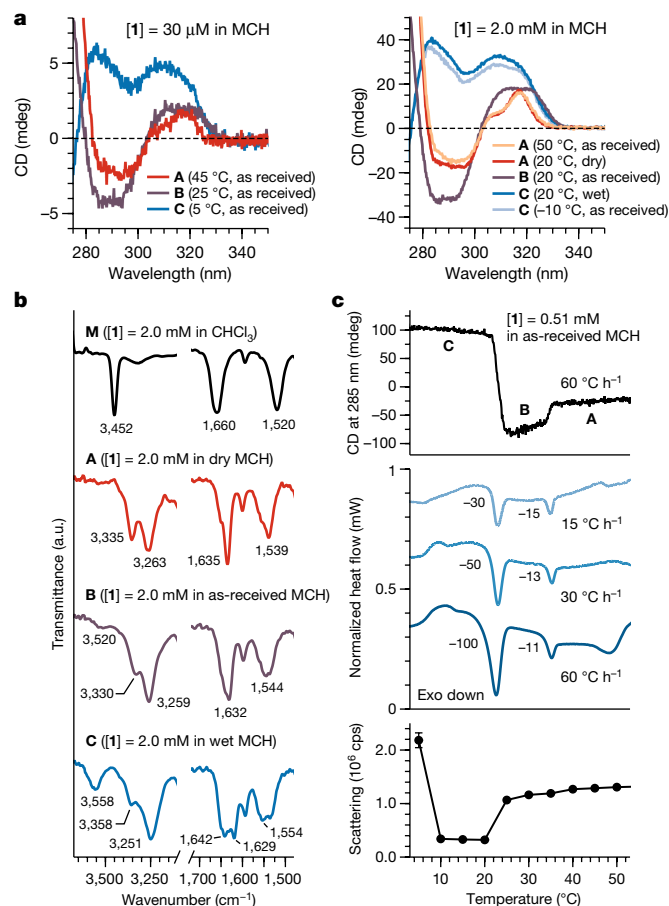
To further establish that water molecules bind to **B** and **C**, we studied each aggregate by Fourier-transform infrared (FTIR) spectroscopy. **A**, **B** and **C** were all produced at 20 °C ([**1**] = 2.0 mM) simply by modulating the water content of the solution (Fig. 2a). For instance, **B** forms at 20 °C when **1** is polymerized in MCH that is used as received; however, when **1** is polymerized in MCH that has been dried with activated molecular sieves, **A** forms instead at 20 °C. FTIR spectroscopy reveals that **A**, **B** and **C** all exhibit red-shifted N–H stretching, red-shifted amide I bands and blue-shifted amide II bands compared to molecularly dissolved **1** in chloroform, typical of hydrogen-bond-based aggregation of carboxamides (Fig. 2b). Of particular interest is the fact that the FTIR spectrum of **B** features a shoulder at 3,520 cm<sup>–1</sup> that is not observed for **A**, and this shoulder appears to evolve into a new band at 3,558 cm<sup>–1</sup> in the

<sup>1</sup>Institute for Complex Molecular Systems, Laboratory of Macromolecular and Organic Chemistry, Eindhoven University of Technology, Eindhoven, The Netherlands. <sup>2</sup>Laboratory of Self-Organizing Soft Matter, Eindhoven University of Technology, Eindhoven, The Netherlands. <sup>3</sup>Crystal and Structural Chemistry, Bijvoet Center for Biomolecular Research, Utrecht University, Utrecht, The Netherlands. <sup>4</sup>Inorganic Materials Chemistry, Eindhoven University of Technology, Eindhoven, The Netherlands. \*e-mail: e.w.meijer@tue.nl



**Fig. 1 | Self-assembly of 1 in MCH.** **a**, Chemical structure of enantiopure 1. **b**, Super-resolution fluorescence microscopy image of a fibre of 1. **c**, AFM image of fibres of 1. **d**, CD and ultraviolet spectra of molecularly dissolved 1 (M), A, B and C while cooling a solution of 1 in MCH ( $[1] = 30 \mu\text{M}$ ,  $[\text{H}_2\text{O}] = 35 \pm 2 \text{ p.p.m.}$ ). **e**, VT-CD cooling experiments in which the concentration of 1 was varied while the water content was held constant ( $[1] = 59 \mu\text{M}$  (darkest shade),  $49 \mu\text{M}$ ,  $40 \mu\text{M}$ ,  $33 \mu\text{M}$ ,  $20 \mu\text{M}$  and  $10 \mu\text{M}$  (lightest shade);  $[\text{H}_2\text{O}] = 35 \pm 2 \text{ p.p.m.}$ ). **f**, VT-CD cooling experiments in which the water content was varied while the concentration of 1 was held constant ( $[\text{H}_2\text{O}] = 47 \pm 3 \text{ p.p.m.}$  (darkest shade),  $35 \pm 2 \text{ p.p.m.}$ ,  $24 \pm 4 \text{ p.p.m.}$ ,  $8 \pm 1 \text{ p.p.m.}$  (lightest shade);  $[1] = 30 \mu\text{M}$ ). All water content measurements are reported as mean  $\pm$  s.d. ( $n = 2$ ).

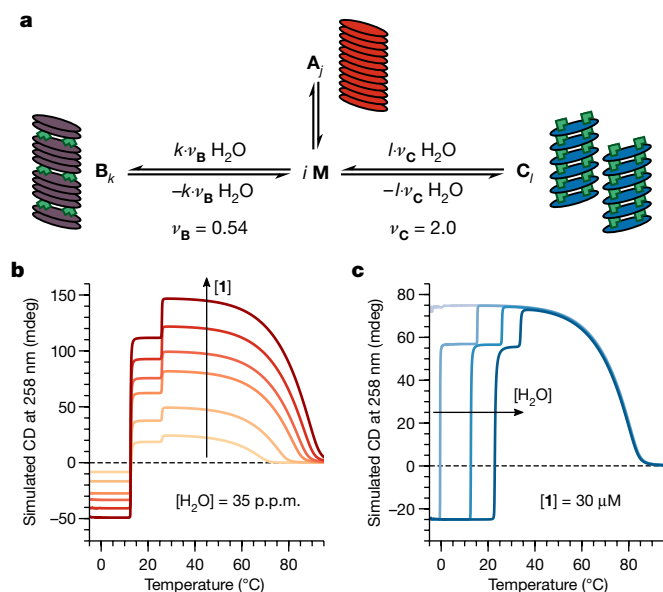
spectrum of C. On the basis of previous studies<sup>19,20</sup>, the frequencies of these bands are consistent with O–H stretching vibrations of water molecules that are engaged in hydrogen bonding through both of their OH groups. Considered together with the accompanying shifts of the N–H stretching band and those of the amide I and amide II bands exhibited by B and C, these measurements indicate that these structures are formed through the incorporation of water.



**Fig. 2 | Characterization of aggregates of 1 by CD spectroscopy, FTIR spectroscopy, micro-DSC and light scattering.** **a**, Diagnostic region of the CD spectra of A, B and C, with the concentration of 1 at  $30 \mu\text{M}$  (left,  $l = 10 \text{ mm}$ ) and  $2.0 \text{ mM}$  (right,  $l = 1 \text{ mm}$ ). **b**, Comparison of the O–H and N–H stretching (left) and the amide I and amide II (right) regions of the FTIR spectra of M, A, B and C (labels in  $\text{cm}^{-1}$ ). **c**, CD signal (top), micro-DSC curves (middle, labels in  $\text{kJ (mol 1)}^{-1}$ ) and light-scattering counts (bottom, mean  $\pm$  s.d. ( $n = 5$ ) are shown) acquired while cooling a solution of 1 in MCH.

If there is a net increase in the number of hydrogen bonds formed, A  $\rightarrow$  B and B  $\rightarrow$  C are expected to evolve heat assuming minimal thermal contributions from changes in solvation. We characterized a  $0.51 \text{ mM}$  solution of 1 by both VT-CD spectroscopy and micro-differential scanning calorimetry (micro-DSC) between  $60^\circ\text{C}$  and  $0^\circ\text{C}$  (Fig. 2c, top and middle) and detected exothermic transitions that clearly correspond to A  $\rightarrow$  B and B  $\rightarrow$  C, respectively, as well as a broad endothermic transition that occurs just below the critical temperature of B  $\rightarrow$  C. Whereas the enthalpy of the A  $\rightarrow$  B transition is insensitive to the scan rate, that of the B  $\rightarrow$  C transition becomes less exothermic from  $-100$  to  $-30 \text{ kJ (mol 1)}^{-1}$  as the scan rate is decreased from  $60$  to  $15^\circ\text{C h}^{-1}$ , which appears to result from the endothermic transition concomitantly shifting to higher temperatures and overlapping more with B  $\rightarrow$  C. We discovered that C scatters less light than B just below the critical temperature of B  $\rightarrow$  C (Fig. 2c, bottom), which suggests that the endothermic transition corresponds to fibre fragmentation reactions. These processes are all reversible, although hysteresis is observed between B  $\rightarrow$  C and C  $\rightarrow$  B (Extended Data Figs. 5, 6, respectively), which is indicative of pathway complexity<sup>21</sup>.

An important question concerns how much water is bound to B and C. To answer this question, we constructed a thermodynamic model consisting of three competing, cooperative supramolecular polymerization pathways<sup>22</sup> (Fig. 3a; see Supplementary Information for details) that enables the derivation of equations for the free energy of elongation of B and C (Supplementary Information equations (13), (18)). Using



**Fig. 3 | Thermodynamic model for the formation of A, B and C.**

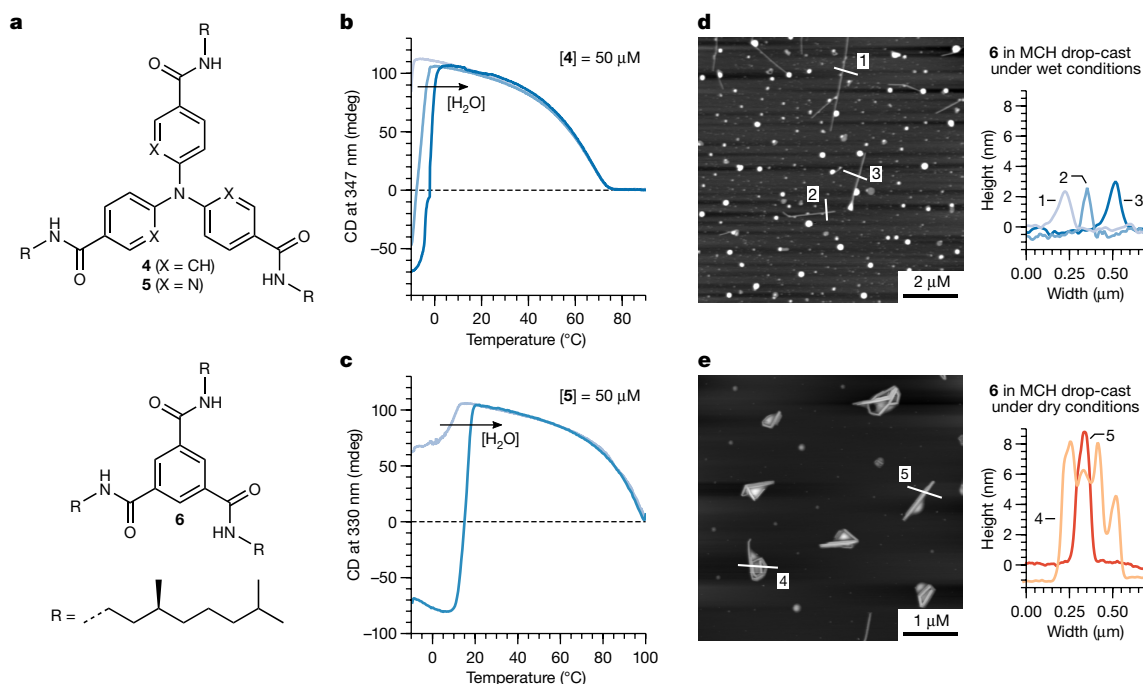
**a**, Schematic representation of three cooperative, competitive pathways. The variables  $j$ ,  $k$  and  $l$  correspond to the degrees of polymerization of **A**, **B** and **C**, respectively. The coloured discs represent aggregated monomer units, and the green blocks represent water molecules. **b**, Simulated VT-CD cooling curves in which the concentration of **I** was varied while the water content was held constant ( $[I] = 59 \mu\text{M}$  (darkest shade),  $49 \mu\text{M}$ ,  $40 \mu\text{M}$ ,  $33 \mu\text{M}$ ,  $20 \mu\text{M}$  and  $10 \mu\text{M}$  (lightest shade);  $[H_2O] = 35$  p.p.m.). **c**, Simulated VT-CD cooling curves in which the water content was varied while the concentration of **I** was held constant ( $[H_2O] = 47$  p.p.m. (darkest shade),  $35$  p.p.m.,  $24$  p.p.m. and  $8$  p.p.m. (lightest shade);  $[I] = 30 \mu\text{M}$ ).

the enthalpy of elongation of **A** (Extended Data Fig. 4a) and the calorimetric enthalpies of  $A \rightarrow B$  and  $B \rightarrow C$  (Fig. 2c), as well as the transition temperatures of  $A \rightarrow B$  and  $B \rightarrow C$  at different concentrations of water

(Fig. 1f), these equations are used to calculate stoichiometries of water for **B** and **C** (that is,  $\nu_B$  and  $\nu_C$  in Fig. 3a) of 0.54 and 2.0, respectively. These calculations also yield values for the entropies of elongation of **B** and **C**; both values agree with those reported for related supramolecular polymerizations<sup>18,23</sup>. Incorporation of these parameters into a temperature-dependent version of the model enables us to simulate the experimental VT-CD curves presented in Fig. 1e, f with a high level of accuracy (Fig. 3b, c, respectively). Although the molecular structures of **B** and **C** remain elusive, we used density functional theory to construct models based on the geometrical parameters obtained from a crystal structure of a chemical analogue of **1** (Extended Data Fig. 7). These models confirm that the incorporation of several water molecules into one-dimensional aggregates of biphenyl tetracarboxamide molecules results in a stable structure<sup>24,25</sup> (Extended Data Fig. 8).

The action of water in oils that is demonstrated here is a result of the potential enthalpic energy of molecularly dissolved water, which is an often overlooked manifestation of hydrophobic effects<sup>26,27</sup>. Although hints that point towards these underlying effects have existed for decades<sup>14,15,28,29</sup>, the present results illustrate the profound molecular consequences of binding even a minuscule amount of water, which prompted us to investigate the effect of water content on the structure of other supramolecular aggregates in oils. We chose to re-examine the self-assembly of triarylamines **4** and **5** (Fig. 4a, top). Both form helical aggregates that exhibit changes in helicity as a function of temperature<sup>30</sup>. VT-CD spectroscopic measurements reveal that the helicities of the resulting aggregates in MCH are sensitive to the concentration of water (Fig. 4b, c), which suggests that water molecules directly interact with these structures as well.

We also investigated the effect of water content on the self-assembly of chiral benzene tricarboxamide **6** (Fig. 4a, bottom), which has been the subject of many previous studies<sup>11</sup>. Although variation of the water content does not affect their helicity, aggregates of **6** prepared in wet MCH scatter more light than fibres formed in dry MCH (Extended Data Fig. 9a). Measurements by AFM suggest an important role for water in modulating the lateral interaction between fibres based on **6**.



**Fig. 4 | Influence of water on the self-assembly of 4, 5 and 6 in MCH.** **a**, Chemical structures of enantiopure **4**, **5** and **6**. **b**, VT-CD cooling experiments in which the water content was varied while the concentration of **4** was held constant ( $[H_2O] = 60 \pm 4$  p.p.m. (darkest shade),  $20 \pm 4$  p.p.m. and  $5 \pm 3$  p.p.m. (lightest shade);  $[4] = 50 \mu\text{M}$ ; cooling rate =  $60^{\circ}\text{C h}^{-1}$ ). **c**, VT-CD cooling experiments in which the water content was varied while the concentration of **5** was held constant

( $[H_2O] = 29 \pm 2$  p.p.m. (darker shade) and  $17 \pm 2$  p.p.m. (lighter shade);  $[5] = 50 \mu\text{M}$ ; cooling rate =  $15^{\circ}\text{C h}^{-1}$ ). All water content measurements in **b** and **c** are reported as mean  $\pm$  s.d. ( $n = 2$ ). **d**, AFM image (left) and height profiles (right) of a sample of **6** prepared in a water-saturated environment. **e**, AFM image (left) and height profiles (right) of a sample of **6** prepared in a glovebox. The difference in morphology observed for aggregates of **6** presented in **d** and **e** was duplicated in our laboratory.

Single-chain fibres are observed when **6** is drop-cast under wet conditions (Fig. 4d and Extended Data Fig. 9b), whereas coils of several chains are detected in samples prepared under dry conditions (Fig. 4e and Extended Data Fig. 9c). We propose that many other unidentified structural transitions observed in hydrogen-bond-based aggregation in oils arise from fundamental interactions with monomeric water molecules in these highly apolar solvents.

### Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0169-0>.

Received: 16 July 2017; Accepted: 16 March 2018;

Published online: 30 May 2018

- Nagornova, N. S., Rizzo, T. R. & Boyarkin, O. V. Interplay of intra- and intermolecular H-bonding in a progressively solvated macrocyclic peptide. *Science* **336**, 320–323 (2012).
- Umena, Y., Kawakami, K., Shen, J. R. & Kamiya, N. Crystal structure of oxygen-evolving photosystem II at a resolution of 1.9 Å. *Nature* **473**, 55–60 (2011).
- Berl, V., Huc, I., Khoury, R. G., Krische, M. J. & Lehn, J. M. Interconversion of single and double helices formed from synthetic molecular strands. *Nature* **407**, 720–723 (2000).
- Hirschberg, J. H. K. K. et al. Helical self-assembled polymers from cooperative stacking of hydrogen-bonded pairs. *Nature* **407**, 167–170 (2000).
- Hartgerink, J. D., Beniash, E. & Stupp, S. I. Self-assembly and mineralization of peptide-amphiphile nanofibers. *Science* **294**, 1684–1688 (2001).
- Jain, S. & Bates, F. S. On the origins of morphological complexity in block copolymer surfactants. *Science* **300**, 460–464 (2003).
- Johnson, R. S., Yamazaki, T., Kovalenko, A. & Fenniri, H. Molecular basis for water-promoted supramolecular chirality inversion in helical rosette nanotubes. *J. Am. Chem. Soc.* **129**, 5735–5743 (2007).
- Appel, E. A. et al. Supramolecular cross-linked networks via host-guest complexation with cucurbit[8]uril. *J. Am. Chem. Soc.* **132**, 14251–14260 (2010).
- Arazoe, H. et al. An autonomous actuator driven by fluctuations in ambient humidity. *Nat. Mater.* **15**, 1084–1089 (2016).
- Aida, T., Meijer, E. W. & Stupp, S. I. Functional supramolecular polymers. *Science* **335**, 813–817 (2012).
- De Greef, T. F. et al. Supramolecular polymerization. *Chem. Rev.* **109**, 5687–5754 (2009).
- Liu, M., Zhang, L. & Wang, T. Supramolecular chirality in self-assembled systems. *Chem. Rev.* **115**, 7304–7397 (2015).
- Riddick, J. A., Bunger, W. B. & Sakano, T. K. *Organic Solvents, Physical Properties and Methods of Purification*. (John Wiley & Sons, New York, 1986).
- Christian, S. D., Taha, A. A. & Gash, B. W. Molecular complexes of water in organic solvents and in the vapour phase. *Q. Rev. Chem. Soc.* **24**, 20–36 (1970).
- Wolfenden, R. & Radzicka, A. On the probability of finding a water molecule in a nonpolar cavity. *Science* **265**, 936–937 (1994).
- Fitié, C. F. et al. Polar switching in trialkylbenzene-1,3,5-tricarboxamides. *J. Phys. Chem. B* **116**, 3928–3937 (2012).
- Smulders, M. M., Schenning, A. P. & Meijer, E. W. Insight into the mechanisms of cooperative self-assembly: the “sergeants-and-soldiers” principle of chiral and achiral C<sub>3</sub>-symmetrical discotic triamides. *J. Am. Chem. Soc.* **130**, 606–611 (2008).
- García, F. et al. The influence of  $\pi$ -conjugated moieties on the thermodynamics of cooperatively self-assembling tricarboxamides. *Chem. Commun.* **49**, 8674–8676 (2013).
- Mohr, S. C., Wilk, W. D. & Barrow, G. M. The association of water with bases and anions in an inert solvent. *J. Am. Chem. Soc.* **87**, 3048–3052 (1965).
- Iwamoto, R. Infrared and near-infrared study of the interaction of amide C=O with water in ideally inert medium. *J. Phys. Chem. A* **114**, 7398–7407 (2010).
- Sorrenti, A., Leira-Iglesias, J., Markvoort, A. J., de Greef, T. F. A. & Hermans, T. M. Non-equilibrium supramolecular polymerization. *Chem. Soc. Rev.* **46**, 5476–5490 (2017).
- Zhao, D. & Moore, J. S. Nucleation-elongation: a mechanism for cooperative supramolecular polymerization. *Org. Biomol. Chem.* **1**, 3471–3491 (2003).
- Venkata Rao, K., Miyajima, D., Nihonyanagi, A. & Aida, T. Thermally bisignate supramolecular polymerization. *Nat. Chem.* **9**, 1133–1139 (2017).
- Gilli, G., Bellucci, F., Ferretti, V. & Bertolasi, V. Evidence for resonance-assisted hydrogen bonding from crystal-structure correlations on the enol form of the  $\beta$ -diketone fragment. *J. Am. Chem. Soc.* **111**, 1023–1028 (1989).
- Filot, I. A. W. et al. Understanding cooperativity in hydrogen-bond-induced supramolecular polymerization: a density functional theory study. *J. Phys. Chem. B* **114**, 13667–13674 (2010).
- Chandler, D. Interfaces and the driving force of hydrophobic assembly. *Nature* **437**, 640–647 (2005).
- Biedermann, F., Nau, W. M. & Schneider, H. J. The hydrophobic effect revisited—studies with supramolecular complexes imply high-energy water as a noncovalent driving force. *Angew. Chem. Int. Ed.* **53**, 11158–11171 (2014).
- Lassette, E. N. The hydrogen bond and association. *Chem. Rev.* **20**, 259–303 (1937).
- Louhichi, A., Jacob, A. R., Bouteiller, L. & Vlassopoulos, D. Humidity affects the viscoelastic properties of supramolecular living polymers. *J. Rheol.* **61**, 1173–1182 (2017).
- Adelizzi, B., Filot, I. A. W., Palmans, A. R. A. & Meijer, E. W. Unravelling the pathway complexity in conformationally flexible N-centered triarylamines trisamides. *Chem. Eur. J.* **23**, 6103–6110 (2017).

**Acknowledgements** This project received funding from the European Union's Horizon 2020 research and innovation program (no. 705701) and the Dutch Ministry of Education, Culture and Science's Gravitation program (no. 024.001.035). The X-ray diffractometer was financed by the Netherlands Organization for Scientific Research. N.J.V.Z. acknowledges M. L. Ślęczkowski for synthesizing the chiral amine precursor for **1**, B. F. M. de Waal for assistance with Karl Fischer titrations and R. A. A. Bovee for performing MALDI-TOF MS measurements. We also thank S. C. J. Meskers, R. J. M. Nolte and A. J. Markvoort for discussions.

**Reviewer information** Nature thanks C. Hunter, D. Miyajima and D. Pantos for their contribution to the peer review of this work.

**Author contributions** N.J.V.Z. conceived the project and designed the experiments. N.J.V.Z., B.A., X.M., R.H.Z. and A.A. performed the experiments and analysed the data. M.F.J.M. and I.A.W.F. performed the mathematical simulations and density functional theory calculations, respectively. M.L. determined the crystal structure of **2**. N.J.V.Z., M.F.J.M., A.A. and I.A.W.F. wrote the manuscript. A.R.A.P. and E.W.M. supervised the research.

**Competing interests** The authors declare no competing interests.

### Additional information

**Extended data** is available for this paper at <https://doi.org/10.1038/s41586-018-0169-0>.

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41586-018-0169-0>.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>.

**Correspondence and requests for materials** should be addressed to E.W.M.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## METHODS

**Materials.** Solvents used for synthesis were acquired from commercial sources and used as received unless otherwise stated. Spectrophotometric grade methylcyclohexane and heptane were purchased from Sigma Aldrich. [1,1'-Biphenyl]-3,3',5,5'-tetracarboxylic acid was either purchased from TCI Europe (>98%) or prepared as previously reported<sup>31</sup>. Pentafluorophenyl trifluoroacetate (98%), 2-methoxyethylamine (99%) and triethylamine (≥99%) were purchased from Sigma Aldrich and used as received. (S)-3,7-Dimethyloctan-1-amine was prepared as previously reported<sup>32</sup> from (S)-citronellol (min. 99%, enantiomeric excess 98.4%) that was purchased from Takasago. Compounds 4,4',4''-nitritoltris(*N*-((S)-3,7-dimethyloctyl)benzamide) (4), 6,6',6''-nitritoltris(*N*-((S)-3,7-dimethyloctyl)nicotinamide) (5) and *N,N',N''*-tris((S)-3,7-dimethyloctyl)benzene-1,3,5-tricarboxamide (6) were prepared as previously reported<sup>29,33</sup>. The photoactivatable dye Cage-552 used for super-resolution fluorescence microscopy was purchased from Abberior.

**Instrumental methods.** Flash column chromatography was performed on a Biotage Isolera One system equipped with an ultraviolet detector. <sup>1</sup>H (400 MHz), <sup>13</sup>C (100 MHz) and <sup>19</sup>F (376 MHz) nuclear magnetic resonance (NMR) spectra were recorded using a Bruker Avance III HD NanoBay spectrometer. <sup>1</sup>H and <sup>13</sup>C NMR spectra were referenced to the residual chloroform signals at 7.26 p.p.m. and 77.23 p.p.m., respectively. NMR spectra are provided in the Supplementary Information. Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) was performed on a Bruker Autoflex Speed spectrometer using α-cyano-4-hydroxycinnamic acid and 2-[(2*E*)-3-(4-*tert*-butylphenyl)-2-methylprop-2-enylidene]malononitrile as matrices. Infrared spectra were acquired using a Perkin Elmer Spectrum Two spectrometer. Solid samples were analysed using a UATR module. Solutions were analysed using a liquid cell (KBr, *l* = 0.5 mm) and a slide holder module. Baseline spectra of as-received MCH were subtracted from sample spectra. DSC was performed on a TA Instruments Q2000 system. About 5 mg of material was prepared in hermetically sealed aluminium pans and characterized using the following heating program: equilibrate at 35 °C, 35 °C to 300 °C at 10 °C min<sup>-1</sup>, 300 °C to -50 °C at -10 °C min<sup>-1</sup>, -50 °C to 300 °C at 10 °C min<sup>-1</sup>. Micro-DSC was performed on a TA Instruments Multicell DSC. About 1.0 ml of material was prepared in Hastelloy ampoules and characterized using the following heating program: equilibrate at 25 °C, 25 °C to 60 °C at 60 °C h<sup>-1</sup>, equilibrate for 15 min, 60 °C to 0 °C at 60 °C h<sup>-1</sup>, equilibrate for 15 min, 0 °C to 60 °C at 60 °C h<sup>-1</sup>, equilibrate for 15 min, 60 °C to 0 °C at 30 °C h<sup>-1</sup>, equilibrate for 15 min, 0 °C to 60 °C at 30 °C h<sup>-1</sup>, equilibrate for 15 min, 60 °C to 0 °C at 15 °C h<sup>-1</sup>, equilibrate for 15 min, 0 °C to 60 °C at 15 °C h<sup>-1</sup>, 60 °C to 25 °C at 60 °C h<sup>-1</sup>. Baseline curves of as-received MCH were subtracted from sample curves. Polarized optical microscopy was performed on a Leica CTR 6000 microscope equipped with two crossed polarizers, a Linkam hot-stage THMS600 as the sample holder, a Linkam TMS94 controller and a Leica DFC420 C camera. Ultraviolet, CD and linear dichroism spectroscopies were performed using a Jasco J-815 spectrometer equipped with a Jasco PTC-348WI Peltier-type temperature controller. The sample holder was purged with nitrogen at a flow rate of 20 l min<sup>-1</sup>. Quartz cuvettes (Hellma Analytics) with path lengths of 10 mm, 5 mm and 1 mm were used. The cuvette with a path length of 10 mm was equipped with a screw cap that was fitted with a Teflon-lined septum, and those with 5- and 1-mm path lengths were sealed with a Teflon stopper. Before all cooling and heating traces were acquired, samples were equilibrated at 95 °C for 15 min. All measurements were conducted using the sealable 10-mm cuvette unless specifically stated. All variable temperature measurements were performed using a cooling rate of 60 °C h<sup>-1</sup> unless otherwise stated. Karl Fischer titrations were performed using a Mettler-Toledo C30 Coulometric KF Titrator loaded with CombiCoulomat Frit KF reagent (for cells with diaphragm, contains methanol, purchased from Merck). Approximately 1 g of sample was used for a typical single Karl Fischer titration. In a modification of a previously reported procedure<sup>34</sup>, light-scattering measurements were performed using a Malvern μV Zetasizer equipped with an 830-nm laser and a scattering angle of 90°. Samples were prepared with as-received MCH that had been filtered through a 0.45 μm Whatman Anotop 10 syringe filter. Measurements were acquired after equilibrating for 10 min at the desired temperature. Samples for wide-angle X-ray scattering were mounted on V1 grade mica sheets with a thickness of 5–7 μm and measured for 15-min exposures using a SAXSLAB GANESHA system equipped with a GeniX-Cu ultralow divergence source producing X-ray photons with a wavelength of 1.54 Å and a flux of 1 × 10<sup>8</sup> photons per second. Scattering patterns were collected using a Pilatus 300K silicon pixel detector, and the beam centre and the *q* range were calibrated using the diffraction peaks of silver behenate. Conversion of two-dimensional images into one-dimensional spectra was accomplished with SaxsGUI software. Domain spacings were calculated using primary scattering peak positions (*q*\*) and interplanar spacings (*d*\* = 2π/*q*\*). For columnar hexagonal morphologies, the centre-to-centre distance was calculated as 2*d*\*/√3. AFM was performed using an Asylum Research MFP-3D system in non-contact tapping mode. Images were processed using Gwyddion 2.49.

**Synthesis.** Tetrakis(perfluorophenyl)[1,1'-biphenyl]-3,3',5,5'-tetracarboxylate (7). [1,1'-Biphenyl]-3,3',5,5'-tetracarboxylic acid (1.0 g, 3.0 mmol, 1.0 equiv.) was suspended in 50 ml of acetonitrile in a 250-ml flask equipped with a magnetic stir bar. Triethylamine (3.1 ml, 24 mmol, 8.0 equiv.) was added, and the mixture became homogeneous after stirring for approximately 10 min. A solution of pentafluorophenyl trifluoroacetate (3.1 ml, 19 mmol, 6.0 equiv.) in acetonitrile (2 ml) was then added dropwise. After the addition was complete, the headspace was purged with argon and the mixture was stirred at room temperature. After 5 h, the mixture was cooled with an ice bath. The solids were collected on a fine sintered glass frit and washed with cold acetonitrile. After drying in a vacuum oven at 50 °C, 7 (3.0 g, 98%) was isolated as a bright white powder that was sufficiently pure for use in subsequent reactions without further purification. <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>): δ 9.11 (t, 2H), 8.81 (d, 4H). <sup>19</sup>F NMR (376 MHz, CDCl<sub>3</sub>): δ -152.14 (m, 8F), -156.49 (m, 4F), -161.45 (m, 8F).

*N*<sup>3</sup>,*N*<sup>3'</sup>,*N*<sup>5</sup>,*N*<sup>5'</sup>-tetrakis((S)-3,7-dimethyloctyl)-[1,1'-biphenyl]-3,3',5,5'-tetracarboxamide (1). Compound 7 (0.32 g, 0.32 mmol, 1.0 equiv.) was dissolved in 10 ml dry THF in a 50-ml flask equipped with a magnetic stir bar. (S)-3,7-Dimethyloctan-1-amine (0.41 g, 2.6 mmol, 8.0 equiv.) and triethylamine (0.36 ml, 2.6 mmol, 8.0 equiv.) were diluted with 3 ml dry THF and added dropwise to the solution of 7 at room temperature. The headspace was purged with argon, and the mixture was stirred for 16 h at 50 °C. After cooling to room temperature, the solvent was removed by rotary evaporation and the crude solid was dissolved in chloroform. This solution was washed with 1 M NaOH, 1 M HCl and brine. The organic phase was dried with MgSO<sub>4</sub> and filtered. The solvent was removed by rotary evaporation, and the resulting solid was purified by flash column chromatography (eluent, 15% ethyl acetate in chloroform). 1 was isolated as a white waxy solid (0.20 g, 70%) after removing the solvent by rotary evaporation and subsequently drying under vacuum (<100 mTorr) in a desiccator with P<sub>2</sub>O<sub>5</sub>. The material was stored in a desiccator loaded with CaSO<sub>4</sub>. <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>): δ 8.02 (m, 2H), 7.90 (m, 4H), 6.76 (broad t, 6.76), 3.50 (m, 8H), 1.08–1.74 (m, 40H), 0.96 (d, 12H), 0.86 (d, 24). <sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>): δ 166.72, 140.13, 136.02, 128.26, 125.02, 39.49, 38.89, 37.41, 36.93, 31.14, 28.18, 24.90, 22.93, 22.83, 19.76. MALDI-TOF MS: calculated *m/z* for C<sub>56</sub>H<sub>94</sub>N<sub>4</sub>O<sub>4</sub>: 886.73, found: 887.77 ([M + H]<sup>+</sup>), 909.76 ([M + Na]<sup>+</sup>).

*N*<sup>3</sup>,*N*<sup>3'</sup>,*N*<sup>5</sup>,*N*<sup>5'</sup>-tetrakis(2-methoxyethyl)-[1,1'-biphenyl]-3,3',5,5'-tetracarboxamide (2). Compound 7 (0.40 g, 0.40 mmol, 1.0 equiv.) was dissolved in 10 ml dry THF in a 50-ml flask equipped with a magnetic stir bar. 2-Methoxyethylamine (0.24 g, 3.2 mmol, 8.0 equiv.) and triethylamine (0.45 ml, 3.2 mmol, 8.0 equiv.) were diluted with 3 ml dry THF and added dropwise to the solution of 7 at room temperature. The headspace was purged with argon, and the mixture was stirred for 16 h at 50 °C. After cooling to room temperature, the solvent was removed by rotary evaporation, and the crude solid was dissolved in chloroform. This solution was washed with 1 M NaOH, 1 M HCl and brine. The organic phase was dried with MgSO<sub>4</sub> and filtered. The crude 2 was serially recrystallized by dissolving in chloroform with methanol and layering with pentane to yield approximately 50 mg (22%) of colourless needles. Crystals suitable for X-ray diffraction were formed by vapour diffusion of pentane into a solution of chloroform and methanol. <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>): δ 8.05 (m, 2H), 7.83 (m, 4H), 7.52 (t, 4H), 3.62–3.72 (m, 16H), 3.39 (s, 12H). <sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>): δ 166.66, 140.09, 135.44, 128.35, 125.12, 71.19, 58.86, 40.27. MALDI-TOF MS: calculated *m/z* for C<sub>28</sub>H<sub>38</sub>N<sub>4</sub>O<sub>8</sub>: 558.27, found: 559.30 ([M + H]<sup>+</sup>), 581.28 ([M + Na]<sup>+</sup>), 597.25 ([M + K]<sup>+</sup>).

**Preparation of samples with variable water content.** The water content of MCH is highly dependent on the relative humidity of the atmosphere if handled without special precautions. Care must be taken to ensure that samples are completely sealed during spectroscopic measurements that take place in an inert atmosphere. Cuvettes equipped with a screw cap and a Teflon-lined septum were found to be best suited for these measurements. To determine the water content after analysis by CD spectroscopy, dilute samples were directly injected into the Karl Fischer titration instrument after withdrawing from the sealed cuvette by syringe. All Karl Fischer titration measurements were performed in duplicate and expressed as mean ± s.d. unless otherwise stated.

At the ambient humidity in the laboratory in which this research was carried out, as-received MCH contained 32 ± 3 p.p.m. H<sub>2</sub>O (mean ± s.d. of four measurements). MCH was dried by sparging with argon and then storing over activated 3 Å molecular sieves overnight in a sealed bottle. After bringing into a nitrogen-filled glovebox, the MCH was passed through a 0.2-μm Whatman Anotop 10 syringe filter. The typical water content for dry MCH prepared in this way was <0.1 p.p.m. (that is, below the level of detection of the Karl Fischer titration). Dry samples were prepared in a nitrogen-filled glovebox with dry MCH, taking special care to use oven-dried glassware and Teflon-lined caps for vials. Alternatively, in a modification of a previously reported procedure<sup>35</sup>, dry samples may be prepared with as-received MCH in sealable cuvettes by exposure to a nitrogen-purged atmosphere (for example, in the sample holder of a CD spectrometer) for a short time at 20 °C before starting the measurement.

Wet MCH was prepared by layering MCH (around 20 ml) over water (around 1 ml) that was purified with an EMD Milipore Mili-Q Integral Water Purification System. After allowing to set for at least 2 h, wet MCH was carefully withdrawn with a syringe from the top layer without disturbing the bottom water phase. Wet samples were prepared on the benchtop; care must be taken to minimize exposure of the solvent to the atmosphere by quickly sealing the sample vials. Wet MCH transferred in this way contained approximately 60 p.p.m. water. To form **C** at 20 °C in MCH (for example, in preparation for characterization by infrared spectroscopy), a small drop of water may be added to a sample of **A** or **B** that has been prepared using dry or as-received MCH, respectively. Gentle agitation to facilitate mixing (that is, inversion of the vial) resulted in the formation of **C** within seconds.

To prepare wet samples for AFM analysis, solutions of **6** were prepared in wet MCH and then drop-cast onto mica. The sample was left to evaporate in a Petri dish that contained 1 ml of water. The dish was covered and allowed to stand overnight at room temperature before imaging. The sample was not allowed to come into direct contact with the water droplet. To prepare dry samples for AFM analysis, solutions of **6** were prepared in dry MCH and then drop-cast onto mica in the glovebox. The dish was covered and allowed to stand overnight in the glovebox at room temperature before imaging.

**Bulk characterization of 1.** Using DSC, phase-transition temperatures and their corresponding enthalpies were determined from the cooling and second heating traces with cooling and heating rates of 10 °C min<sup>-1</sup> (Extended Data Fig. 1a). The isotropic melt was reached at 273 °C, and another phase transition was observed at 262 °C; the total enthalpy of these two transitions is 38 J g<sup>-1</sup>. The corresponding exothermic transitions were observed in the cooling trace at 269 °C and 259 °C, respectively. No additional phase transitions were observed between 259 °C and -50 °C. Using polarized optical microscopy, mosaic and focal conic textures were observed, similar to those reported for benzene tricarboxamide-based liquid crystals<sup>36</sup>. Extended Data Fig. 1b shows this texture at 135 °C after cooling from the isotropic melt at 10 °C min<sup>-1</sup>. The liquid crystalline structure of **1** was characterized by wide-angle X-ray scattering. The scattering pattern as well as the lattice parameters of **1** at 20 °C are shown in Extended Data Fig. 1c. In the small angle regime, three reflections were observed with  $q$  values of 3.0, 5.2 and 6.0 nm<sup>-1</sup>. The reciprocal spacing ratio is calculated as 1:1.3:2, which is assigned to the columnar hexagonal liquid crystal structure with the corresponding lattice distances  $d[100] = 2.1$  nm,  $d[110] = 1.2$  nm and  $d[200] = 1.1$  nm, respectively. From  $d[100]$ , the domain spacing,  $L_0$ , is 24 Å. In the wide-angle regime, the reflection at 19.8 nm<sup>-1</sup> gives a  $d$ -spacing of 3.2 Å, which we assign to the interdiscotic distance. These dimensions are similar to those observed for benzene tricarboxamide-based aggregates<sup>16</sup>.

Bulk **1** was further characterized by CD and FTIR spectroscopies. Bulk **1** was prepared for analysis by CD spectroscopy by drop-casting a 2.0 mM solution of **1** in MCH onto a quartz slide. After most of the MCH had evaporated in the ambient atmosphere, the film was heated to 100 °C using the Peltier controller of the CD spectrometer for 1 h. The resulting spectrum (Extended Data Fig. 1d) shows a positive Cotton effect at 255 nm. The full FTIR spectrum of bulk **1** cooled from isotropic melt is presented in Extended Data Fig. 1e, and the comparison of the N-H stretching and the amide I and amide II bands of bulk **1** and **A** (**1** = 2.0 mM in dry MCH) are shown in Extended Data Fig. 1f.

**van 't Hoff analyses.** The enthalpy of elongation,  $\Delta H_e$ , for the formation of **A** was estimated from a van 't Hoff plot of  $\ln(K_e)$  versus  $1/T_e$  (Extended Data Fig. 4a;  $K_e = a_1^{-1}$ ,  $a_1 = [1]/[1]_{\text{ref}}$ ,  $[1]_{\text{ref}} = 1 \mu\text{M}$  in MCH and  $T_e$  is the elongation temperature at a given  $[1]$ ). The elongation temperatures were identified using scripts written previously<sup>37</sup>. This procedure has been previously used to estimate the  $\Delta H_e$  for other cooperative supramolecular polymerizations<sup>17</sup>. For **A**  $\rightarrow$  **B** and **B**  $\rightarrow$  **C**, an analogous procedure was used to estimate the corresponding molar enthalpies of hydration for each process,  $\Delta H_{\text{hyd,A}}$  and  $\Delta H_{\text{hyd,B}}$ , respectively (Extended Data Fig. 4b, c). We assumed that **A** and **B** are in equilibrium with molecularly dissolved water in each respective transition. For each transition,  $\ln(K_{\text{hyd,A}})$  and  $\ln(K_{\text{hyd,B}})$  ( $K_{\text{hyd}} = a_{\text{H}_2\text{O}}^{-1}$ ,  $a_{\text{H}_2\text{O}} = [\text{H}_2\text{O}]/[\text{H}_2\text{O}]_{\text{ref}}$ , and  $[\text{H}_2\text{O}]_{\text{ref}} = 1 \mu\text{M}$  in MCH) were plotted against  $1/T_{\text{A} \rightarrow \text{B}}$  and  $1/T_{\text{B} \rightarrow \text{C}}$ , respectively. The critical temperatures  $T_{\text{A} \rightarrow \text{B}}$  and  $T_{\text{B} \rightarrow \text{C}}$  were identified using the second derivative of each corresponding VT-CD curve (Extended Data Fig. 4d). On the basis of the independence of **A**  $\rightarrow$  **B** and **B**  $\rightarrow$  **C** to the concentration of **1** (Fig. 1e), we assumed that the activities of **A**, **B** and **C** were constants.

**Super-resolution microscopy.** A modification of previously reported PALM<sup>38</sup> and iPAINT<sup>39</sup> protocols were used to visualize fibres of **1** and **6**. A detailed description of this technique is in preparation and will be published elsewhere. In short, 200  $\mu\text{M}$  solutions of each aggregate were stained with 1% v/v of Cage-552 (10 mM in DMSO) and 1% v/v of *i*-PrOH. Each prepared sample was injected into a sample chamber constructed from a glass cover slide and coverslip held together with double-sided tape, and iPAINT images were acquired using a Nikon N-STORM system as described previously<sup>39</sup>. Time-lapses of  $15 \times 10^3$  frames were recorded onto a  $256 \times 256$  pixel region (pixel size 170 nm) of an EMCCD camera (ixon3, Andor) at a rate of 47 frames per second. To perform single-molecule experiments, a low

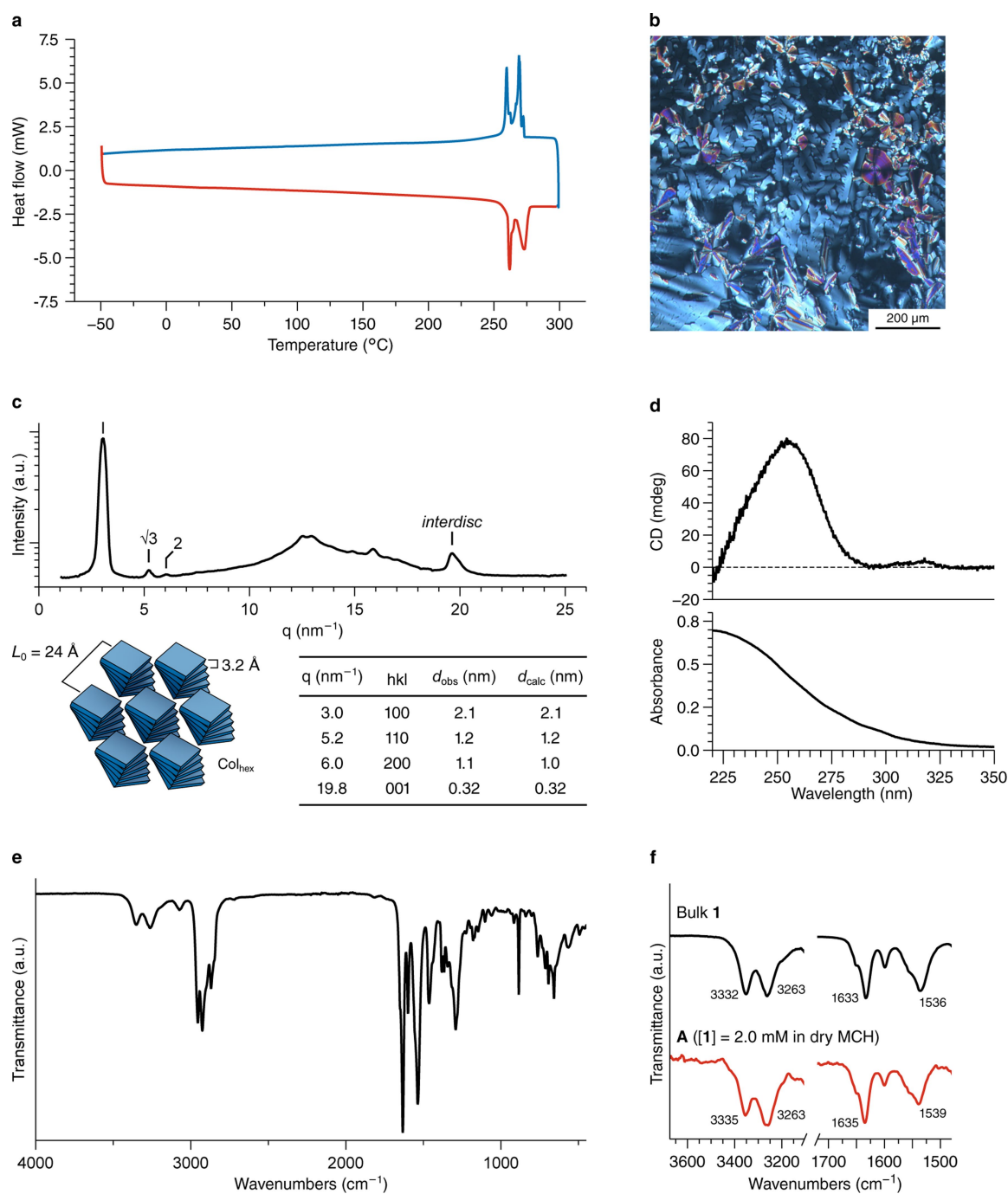
ultraviolet laser light power (1.6 mW cm<sup>-2</sup> power at 405 nm) is used to uncage a small amount of dye per frame, statistically ensuring a spatial separation greater than the diffraction limit of light. The sample is subsequently irradiated at 561 nm with a laser of optical intensity 488 mW cm<sup>-2</sup> to excite the single molecules that were previously photoactivated. The high-power laser bleaches the excited molecules, so that a new subset of molecules can be photoactivated, excited and localized. The localization of single molecules is finally carried out by NIS-element Nikon software. The procedure for the thickness analyses is detailed in the Supplementary Information.

**X-ray crystal structure determination of 2.** C<sub>28</sub>H<sub>38</sub>N<sub>4</sub>O<sub>8</sub>·2CHCl<sub>3</sub>,  $M_w = 797.36$  g mol<sup>-1</sup>, colourless needle,  $0.53 \times 0.14 \times 0.02$  mm<sup>3</sup>, triclinic,  $P\bar{1}$  (no. 2),  $a = 12.2183(9)$ ,  $b = 14.9730(9)$ ,  $c = 16.0255(9)$  Å,  $\alpha = 73.220(3)$ ,  $\beta = 82.014(2)$ ,  $\gamma = 77.416(3)^\circ$ ,  $V = 2,730.4(3)$  Å<sup>3</sup>,  $Z = 3$ ,  $D_x = 1.455$  g cm<sup>-3</sup>,  $\mu = 0.53$  mm<sup>-1</sup>, 70,947 reflections were measured on a Bruker Kappa ApexII diffractometer with sealed tube and Triumph monochromator ( $\lambda = 0.71073$  Å) at a temperature of 150(2) K up to a resolution of  $(\sin \theta/\lambda)_{\text{max}} = 0.65$  Å<sup>-1</sup>. The intensities were integrated with the Eval15 software<sup>40</sup>. Numerical absorption correction and scaling was performed with SADABS<sup>41</sup> (correction range 0.73–1.00). 12,556 reflections were unique ( $R_{\text{int}} = 0.062$ ), of which 7,428 were observed ( $I > 2\sigma(I)$ ). The structure was solved with Patterson superposition methods using SHELXT<sup>42</sup>. Least-squares refinement was performed with SHELXL-2016<sup>43</sup> against  $F^2$  of all reflections. Non-hydrogen atoms were refined freely with anisotropic displacement parameters. One of the chloroform molecules was refined with a disorder model. Hydrogen atoms were introduced in calculated positions. N–H hydrogen atoms were refined freely with isotropic displacement parameters. C–H hydrogen atoms were refined with a riding model. 707 parameters were refined with 168 restraints (distances, angles and displacement parameters of the chloroform molecules).  $R1/wR2$  ( $I > 2\sigma(I)$ ): 0.0619/0.1528.  $R1/wR2$  (all reflections): 0.1189/0.1826.  $S = 1.029$ . Residual electron density between -0.74 and 1.10 e Å<sup>-3</sup>. Geometry calculations and checking for higher symmetry were performed with the PLATON program<sup>44</sup>.

**Computational settings.** Density functional theory simulations were performed using VASP. The PBE exchange-correlation functional was used in conjunction with the projector augmented wave approach. All structures were optimized to their local minima using the conjugate gradient algorithm as implemented in VASP. The nature of the stationary points was evaluated from the harmonic modes, computed numerically by using a complete Hessian matrix (that is, incorporating all degrees of freedom). No imaginary frequencies were found for the optimized structures with the exception of some degrees of freedom corresponding to rotation and translation, which confirms that these geometries correspond to local minima on the potential energy surface. Optimization and other electronic settings are provided in the Supplementary Information.

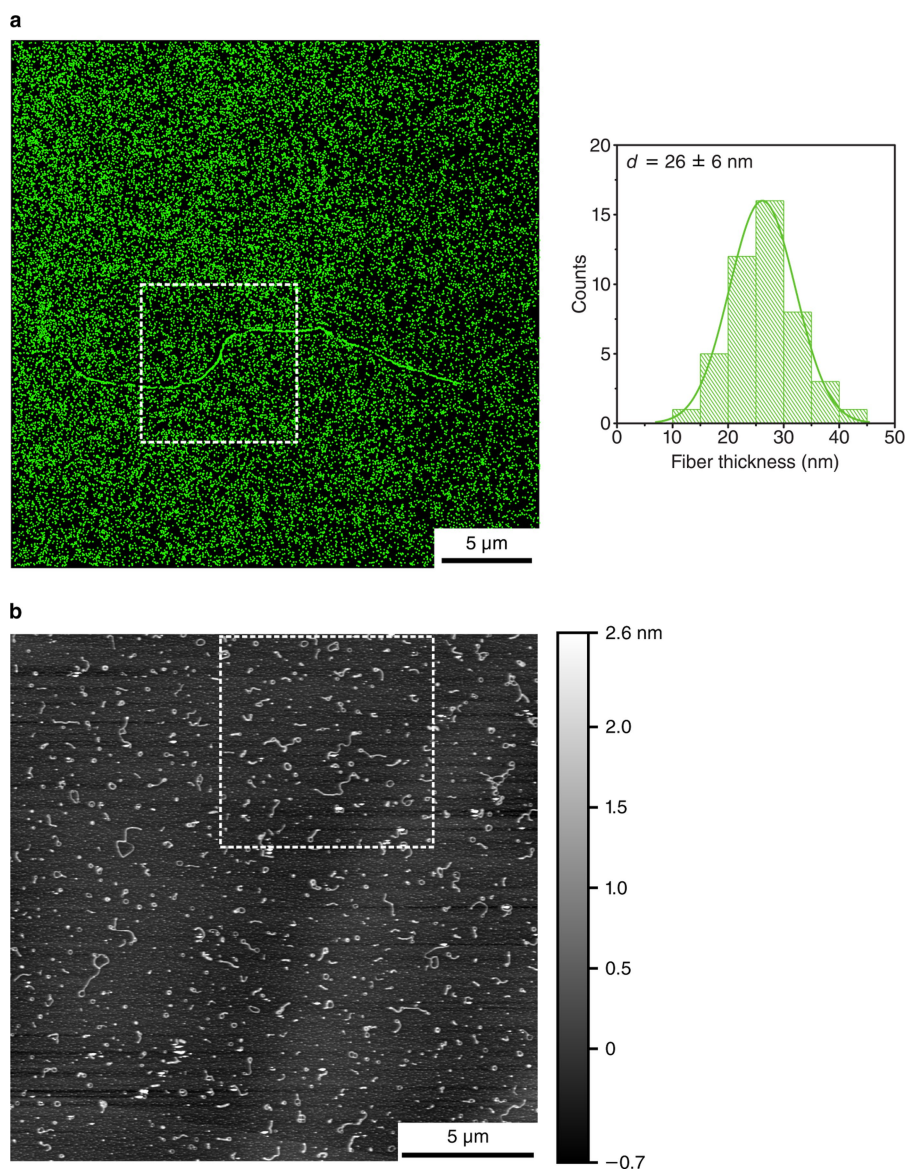
**Data availability.** Data that support the findings of this study are available within the paper and its Supplementary Information. Metrical parameters for **2** are available free of charge from the Cambridge Crystallographic Data Centre (<https://www.ccdc.cam.ac.uk>) under reference number CCDC 1562237.

- Mazik, M. & König, A. Recognition properties of an acyclic biphenyl-based receptor toward carbohydrates. *J. Org. Chem.* **71**, 7854–7857 (2006).
- Suthagar, K., Watson, A. J., Wilkinson, B. L. & Fairbanks, A. J. Synthesis of arabinose glycosyl sulfamides as potential inhibitors of mycobacterial cell wall biosynthesis. *Eur. J. Med. Chem.* **102**, 153–166 (2015).
- Brunsveld, L. et al. Chiral amplification in columns of self-assembled *N,N,N'*-tris((S)-3,7-dimethyloctyl)benzene-1,3,5-tricarboxamide in dilute solution. *Chem. Lett.* **29**, 292–293 (2000).
- Boekhoven, J. et al. Dissipative self-assembly of a molecular gelator by using a chemical fuel. *Angew. Chem. Int. Ed.* **49**, 4825–4828 (2010).
- Christian, S. D., Afsprung, H. E., Johnson, J. R. & Worley, J. D. Control and measurement of water activity. *J. Chem. Educ.* **40**, 419–421 (1963).
- van Gorp, J. J., Vekemans, J. A. J. M. & Meijer, E. W. C<sub>3</sub>-symmetrical supramolecular architectures: fibers and organic gels from discotic trisamides and trisureas. *J. Am. Chem. Soc.* **124**, 14759–14769 (2002).
- Markvoort, A. J., ten Eikelder, H. M., Hilbers, P. A., de Greef, T. F. & Meijer, E. W. Theoretical models of nonlinear effects in two-component cooperative supramolecular copolymerizations. *Nat. Commun.* **2**, 509 (2011).
- Aloi, A. et al. Imaging nanostructures by single-molecule localization microscopy in organic solvents. *J. Am. Chem. Soc.* **138**, 2953–2956 (2016).
- Aloi, A., Vilanova, N., Albertazzi, L. & Voets, I. K. iPAINT: a general approach tailored to image the topology of interfaces with nanometer resolution. *Nanoscale* **8**, 8712–8716 (2016).
- Schreurs, A. M. M., Xian, X. & Kroon-Batenburg, L. M. J. EVAL15: a diffraction data integration method based on ab initio predicted profiles. *J. Appl. Crystallogr.* **43**, 70–82 (2010).
- Sheldrick, G. M. SADABS (Univ. Göttingen, 2008).
- Sheldrick, G. M. SHELXT - integrated space-group and crystal-structure determination. *Acta Crystallogr. A* **71**, 3–8 (2015).
- Sheldrick, G. M. Crystal structure refinement with SHELXL. *Acta Crystallogr. C* **71**, 3–8 (2015).
- Spek, A. L. Structure validation in chemical crystallography. *Acta Crystallogr. D* **65**, 148–155 (2009).



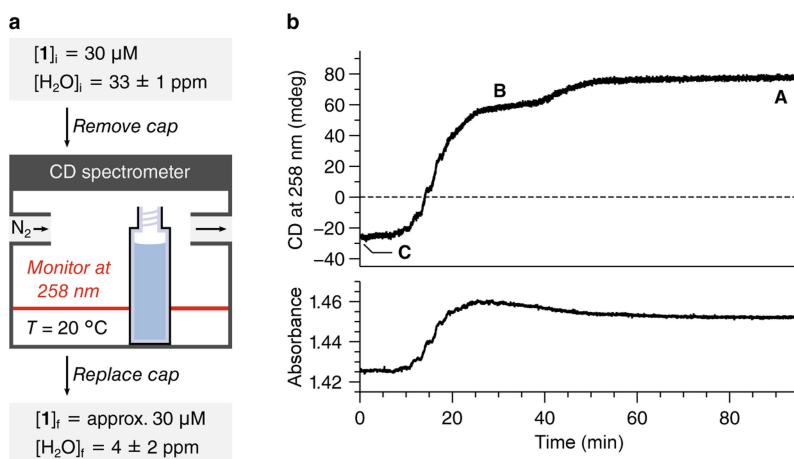
**Extended Data Fig. 1 | Bulk characterization of 1.** **a**, DSC trace of **1** (cooling in blue, second heating in red). **b**, Polarized optical microscopy image of **1** with crossed polarizers at 135 °C after cooling from the isotropic melt. **c**, Wide-angle X-ray scattering trace of bulk **1** at room temperature (top) with a schematic of the hexagonal columnar

morphology (bottom left) and tabulated parameters (bottom right). **d**, CD signal (top) and absorbance (bottom) of a thin film of **1** at 20 °C. **e**, FTIR spectrum of bulk **1** at 20 °C after cooling from the isotropic melt. **f**, Comparison of the FTIR spectra of bulk **1** after cooling from the isotropic melt and **A** ([**1**] = 2.0 mM in dry MCH, labels in  $\text{cm}^{-1}$ ).



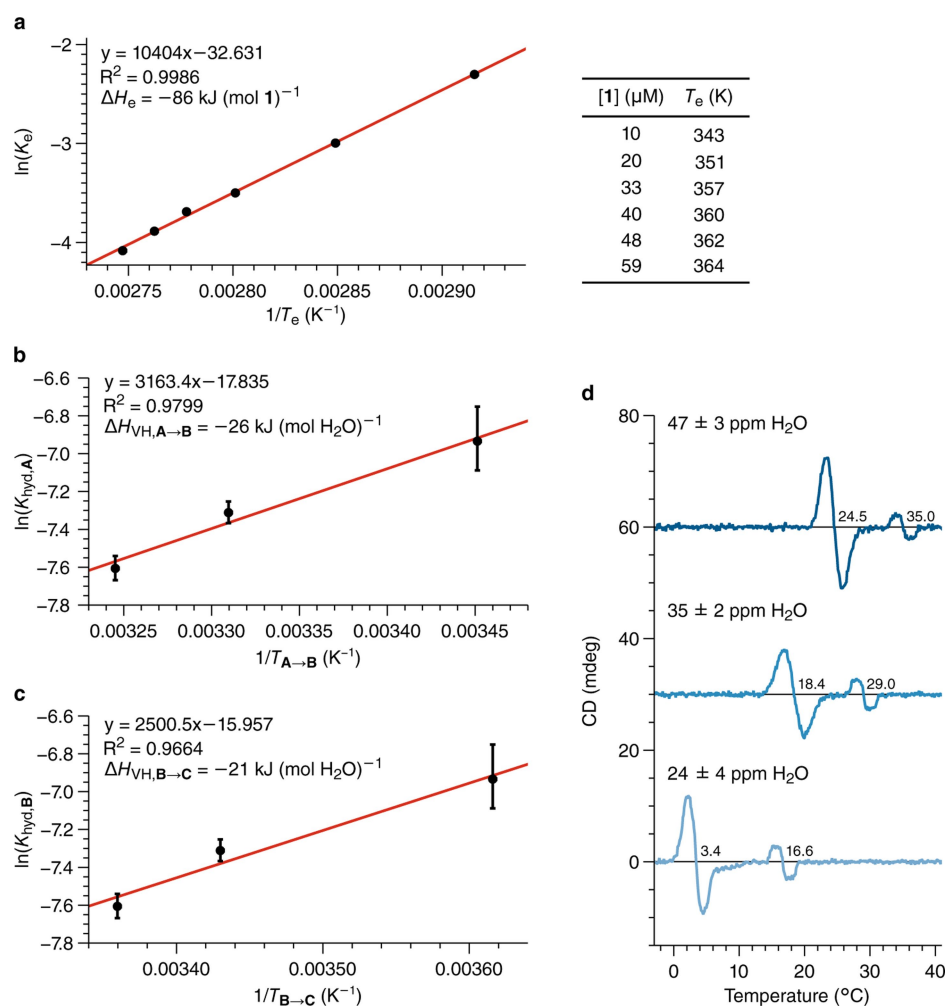
**Extended Data Fig. 2 | Super-resolution fluorescence and AFM images of 1.** **a**, Supramolecular fibres of **1** stained with Cage-552 photoactivatable dye imaged by super-resolution fluorescence microscopy (left) and corresponding thickness analysis (right). The indicated zone in the

microscopy image is depicted in Fig. 1b. A discussion of the thickness analysis and a comparison to fibres of **6** is provided in the Supplementary Information. **b**, Supramolecular fibres of **1** imaged by AFM in non-contact tapping mode. The indicated zone is depicted in Fig. 1c.



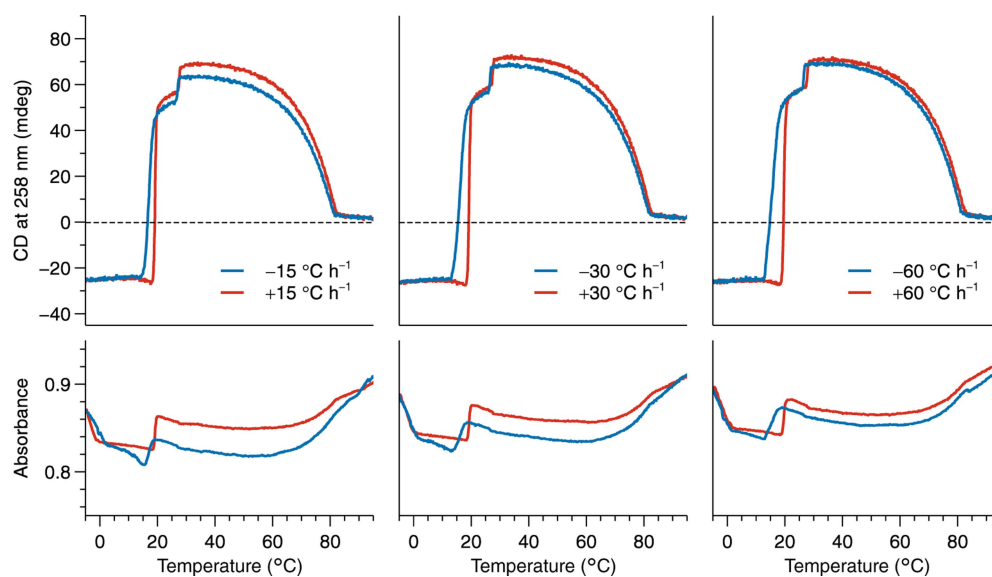
**Extended Data Fig. 3 | Removal of water from aggregates of **1** to effect helicity transitions.** **a**, Schematic of experimental design. The CD spectrometer was purged with nitrogen at a rate of  $20 \text{ l min}^{-1}$ . **b**, CD

signal (top) and absorbance (bottom) at 258 nm as a  $30 \mu\text{M}$  solution of **1** is dried over 100 min in the sample holder of the CD spectrometer. All water content measurements are reported as mean  $\pm$  s.d. ( $n = 2$ ).



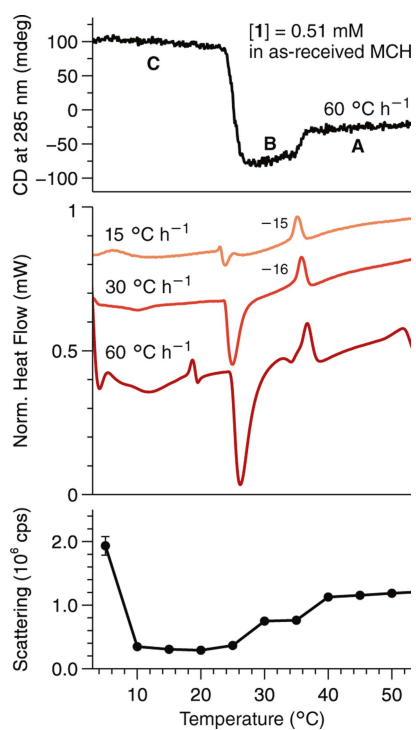
**Extended Data Fig. 4 | van 't Hoff analyses.** **a**, van 't Hoff plot of  $\ln(K_e)$  versus  $1/T_e$  (left) with tabulated  $T_e$  data (right). **b**, van 't Hoff plot of  $\ln(K_{\text{hyd},A})$  versus  $1/T_{A \rightarrow B}$ . The points represent  $\ln(K_{\text{hyd},A})$  calculated using the mean of the water content determined for each respective measurement. The error bars correspond to the spread of  $\ln(K_{\text{hyd},A})$  as a result of the experimental uncertainty of each respective water content measurement. **c**, van 't Hoff plot of  $\ln(K_{\text{hyd},B})$  versus  $1/T_{B \rightarrow C}$ . The points

represent  $\ln(K_{\text{hyd},B})$  calculated using the mean of the water content determined for each respective measurement. The error bars correspond to the spread of  $\ln(K_{\text{hyd},B})$  as a result of the experimental uncertainty of each respective water content measurement. **d**, Determination of  $T_{A \rightarrow B}$  and  $T_{B \rightarrow C}$  from the second derivative of the corresponding VT-CD curves presented in Fig. 1f (labels in °C).



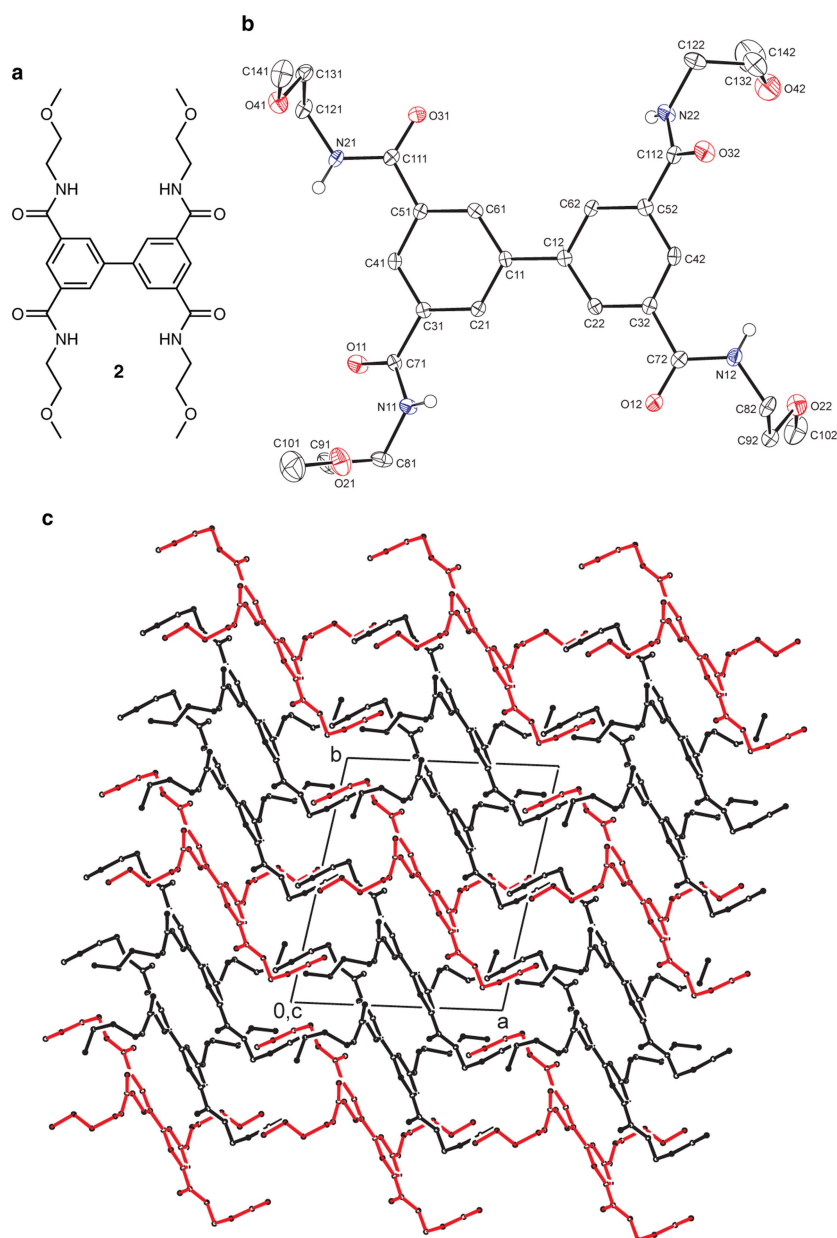
**Extended Data Fig. 5 | Cooling and heating experiments using VT-CD spectroscopy.** The CD intensity was monitored at 258 nm while cooling from 95 °C to −5 °C and then immediately heating back to 95 °C with

scanning rates of 15 (left), 30 (middle) and 60 °C h<sup>−1</sup> (right). Samples were prepared with as-received MCH ([**1**] = 30 μM, [H<sub>2</sub>O] = 35 ± 2 p.p.m.).



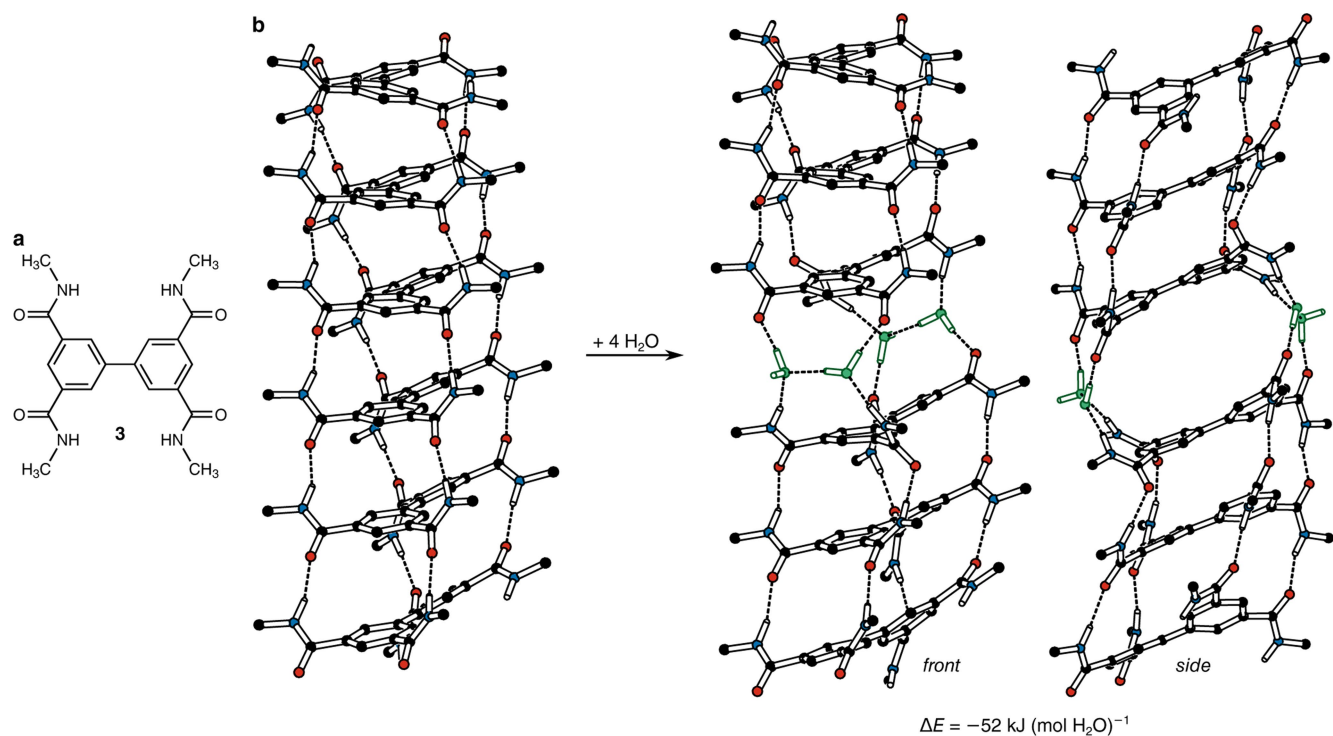
#### Extended Data Fig. 6 | Heating experiments with aggregates of **1**.

A 0.51  $\mu\text{M}$  solution of **1** in as-received MCH was characterized by VT-CD spectroscopy (top), micro-DSC (middle) and light scattering (bottom, mean  $\pm$  s.d. ( $n = 5$ ) are shown). In the micro-DSC plot, only the endothermic transitions corresponding to **B**  $\rightarrow$  **A** with scan rates of 15 and 30  $^{\circ}\text{C h}^{-1}$  had baselines suitable for integration (labels in  $\text{kJ (mol 1)}^{-1}$ ).



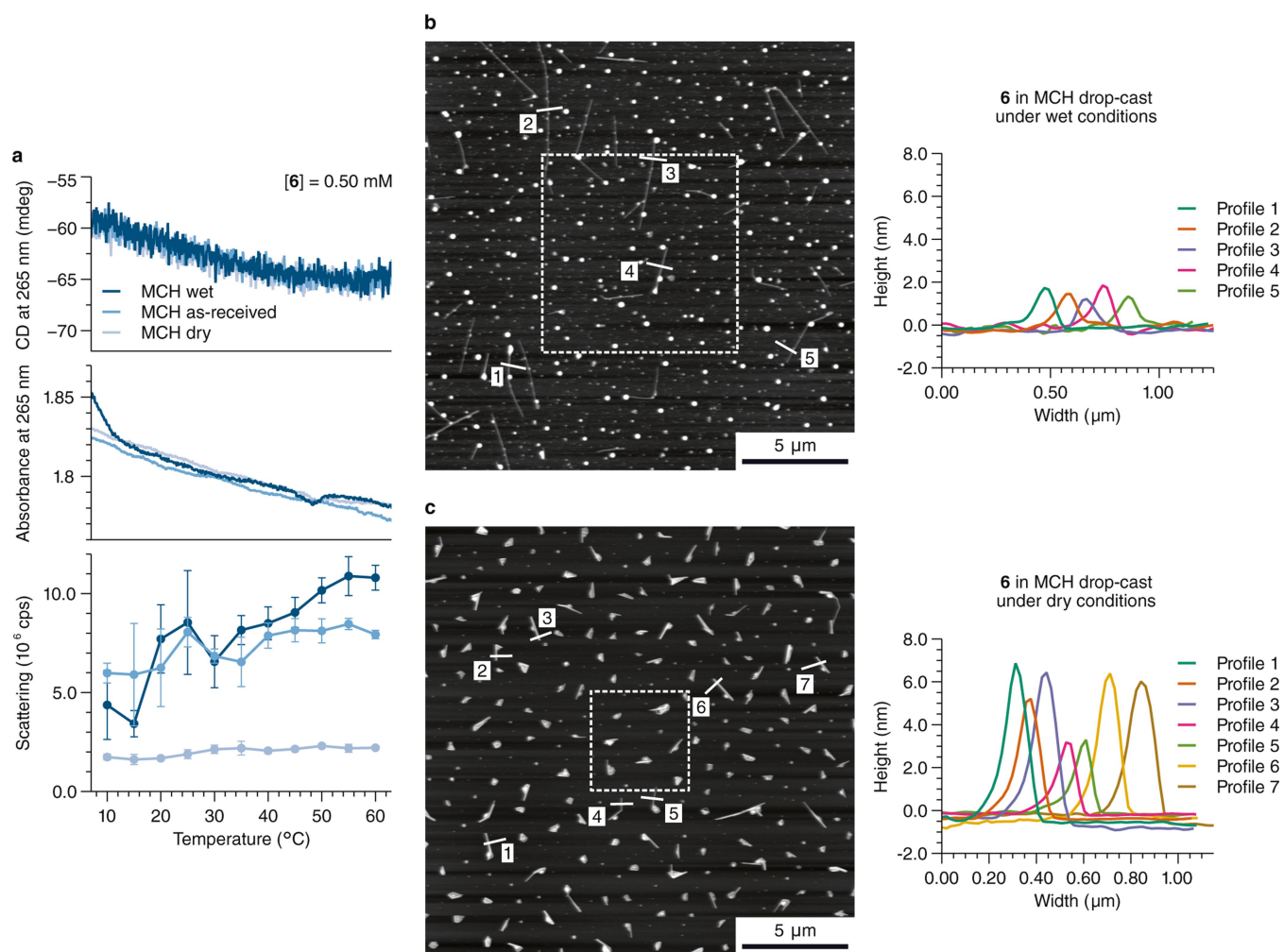
**Extended Data Fig. 7 | Crystal structure of 2.** **a**, Chemical structure of **2**. **b**, Displacement ellipsoid plot (50% probability level) of **2** in the crystal. C–H hydrogen atoms and chloroform solvent molecules are omitted for clarity. Only one of two independent molecules is shown. The other independent molecule is located on an inversion centre. **c**, Packing of **2**

in the crystal. The two independent molecules are shown in black and red, respectively. Hydrogen atoms and chloroform solvent molecules are omitted for clarity. The structure shows pseudo-translational symmetry in the *b*-direction.



**Extended Data Fig. 8 | Molecular model of water binding to an aggregate of biphenyl tetracarboxamide molecules. a,** Chemical structure of **3**. **b,** Molecular models based on density functional theory calculations for the incorporation of four water molecules into a

hexameric aggregate of **3**. Hydrogen atoms, apart from those engaged in hydrogen bonding, are omitted for clarity. The structures are colour coded as follows: hydrogen bond, dashed lines; carbon, black; oxygen, red; nitrogen, blue; water molecules, green.



**Extended Data Fig. 9 | The influence of water content on the self-assembly of 6 in MCH.** **a**, CD signal (top), ultraviolet absorbance (middle) and light-scattering counts (bottom, mean  $\pm$  s.d. ( $n=5$ )) are shown) acquired while cooling solutions of **6** in wet, as-received or dry MCH. **b**, Typical AFM picture (left) and height profiles (right) of a sample of

**6** (30  $\mu\text{M}$  in wet MCH) that was drop-cast on mica in a water-saturated environment. The indicated zone is depicted in Fig. 4d. **c**, Typical AFM picture (left) and height profiles (right) of a sample of **6** (30  $\mu\text{M}$  in dry MCH) that was drop-cast on mica under dry conditions in a glovebox. The indicated zone is depicted in Fig. 4e.

# A global slowdown of tropical-cyclone translation speed

James P. Kossin<sup>1\*</sup>

**As the Earth's atmosphere warms, the atmospheric circulation changes. These changes vary by region and time of year, but there is evidence that anthropogenic warming causes a general weakening of summertime tropical circulation<sup>1–8</sup>. Because tropical cyclones are carried along within their ambient environmental wind, there is a plausible a priori expectation that the translation speed of tropical cyclones has slowed with warming. In addition to circulation changes, anthropogenic warming causes increases in atmospheric water-vapour capacity, which are generally expected to increase precipitation rates<sup>9</sup>. Rain rates near the centres of tropical cyclones are also expected to increase with increasing global temperatures<sup>10–12</sup>. The amount of tropical-cyclone-related rainfall that any given local area will experience is proportional to the rain rates and inversely proportional to the translation speeds of tropical cyclones. Here I show that tropical-cyclone translation speed has decreased globally by 10 per cent over the period 1949–2016, which is very likely to have compounded, and possibly dominated, any increases in local rainfall totals that may have occurred as a result of increased tropical-cyclone rain rates. The magnitude of the slowdown varies substantially by region and by latitude, but is generally consistent with expected changes in atmospheric circulation forced by anthropogenic emissions. Of particular importance is the slowdown of 30 per cent and 20 per cent over land areas affected by western North Pacific and North Atlantic tropical cyclones, respectively, and the slowdown of 19 per cent over land areas in the Australian region. The unprecedented rainfall totals associated with the 'stall' of Hurricane Harvey<sup>13–15</sup> over Texas in 2017 provide a notable example of the relationship between regional rainfall amounts and tropical-cyclone translation speed. Any systematic past or future change in the translation speed of tropical cyclones, particularly over land, is therefore highly relevant when considering potential changes in local rainfall totals.**

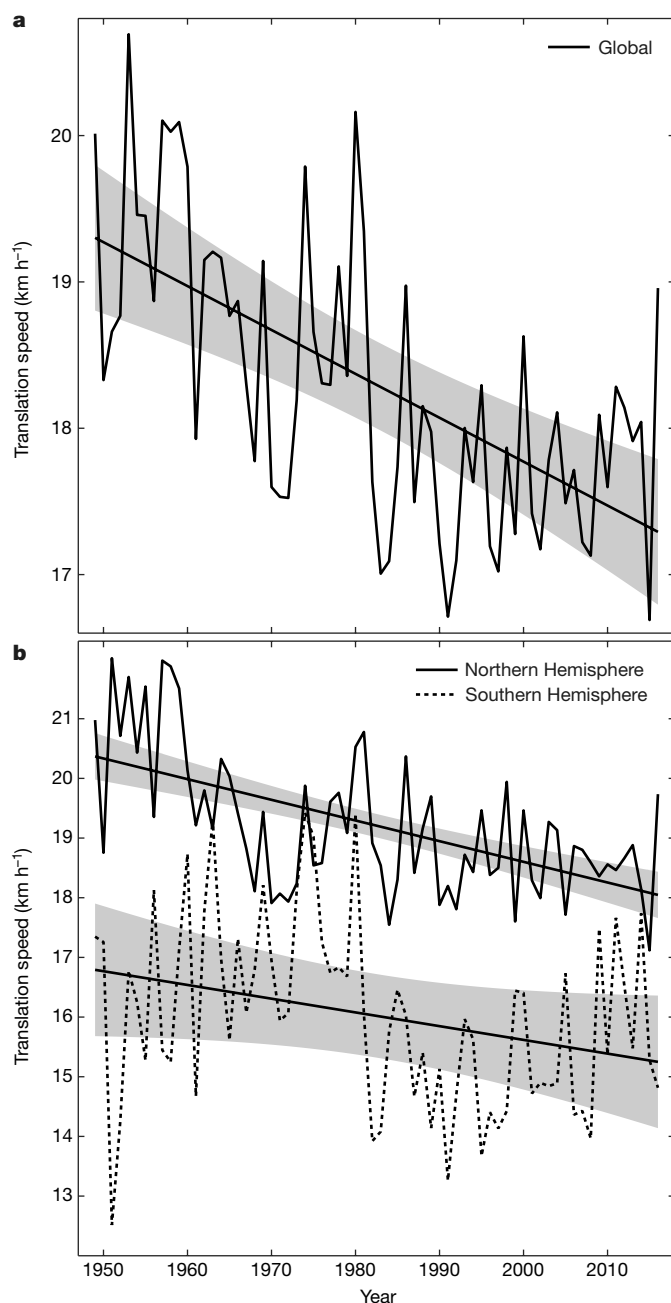
There is complex interaction between the internal and external factors that control tropical-cyclone intensity<sup>16,17</sup>, which is generally described in terms of cyclonic rotation speed. In addition to a rotation speed, tropical cyclones also have a translation speed that is controlled largely by the environmental steering winds in which they are embedded. The maximum tropical-cyclone intensity experienced at ground level is to the right (left) of the translation vector in the Northern (Southern) Hemisphere, where the rotation and translation speeds are summed. Consequently, a slowing of the tropical-cyclone translation speed would reduce the maximum ground-relative intensity in both hemispheres. The ratio of translation to rotation speed can be close to or greater than unity in weaker tropical cyclones, particularly outside of the tropics where the ambient steering winds can be strong. But this ratio is generally small for the more societally relevant, intense tropical cyclones, particularly those translating within the comparatively weak tropical trade winds; in this case, a slowing of translation speed would have a proportionally small effect on the ground-relative maximum intensity. Alternatively, a slowing of translation speed could have a profound effect on the amount of local, tropical-cyclone-related rainfall, which

is proportional to the rate of rain produced in a tropical cyclone and inversely proportional to its translation speed; that is, a proportional unit of decrease in translation speed would have about the same effect on local rainfall totals as the same proportional unit of increase in rain rate.

Anthropogenic warming, both past and projected, is expected to affect the strength and patterns of global atmospheric circulation<sup>1–8</sup>. Tropical cyclones are generally carried along within these circulation patterns so their past translation speeds may be indicative of past circulation changes. In particular, warming is linked to a weakening of tropical summertime circulation and there is a plausible a priori expectation that tropical-cyclone translation speed may be decreasing. In addition to changing circulation, anthropogenic warming is expected to increase lower-tropospheric water-vapour capacity by about 7% per degree (Celsius) of warming, as per the Clausius–Clapeyron relationship<sup>2</sup>. Expectations of increased mean precipitation under global warming are well documented, but not as straightforward to quantify<sup>9,18</sup>. Increases in global precipitation are constrained by the atmospheric energy budget to about 1%–2% per degree of warming<sup>19,20</sup>; those in regional precipitation are further controlled by variability in moisture convergence driven by variability in regional circulation. Precipitation extremes can vary more broadly and are less constrained by energy considerations than is global precipitation<sup>21–24</sup>.

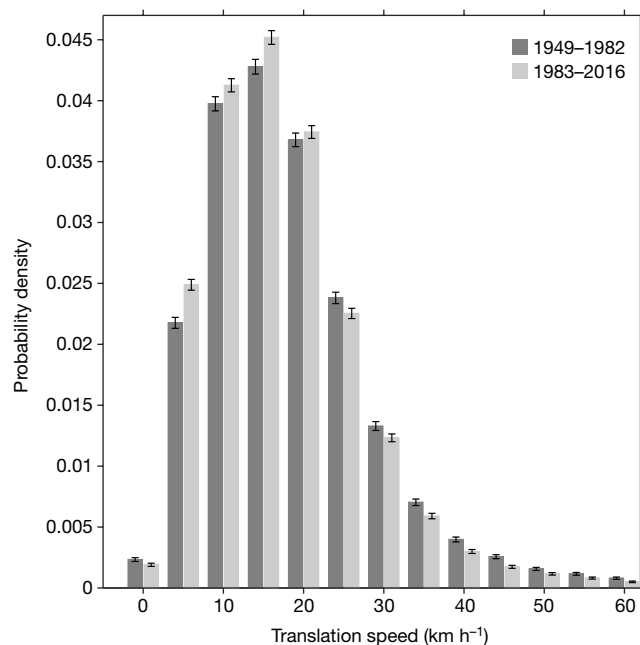
Numerical simulations of tropical-cyclone rain rates are fairly consistent in projecting increases in a warming world<sup>10–12</sup>. Tropical cyclones are very effective at converging moisture, and rain-rate increases tend to be largest near their centres, where the convergence is greatest; further from their centres, simulated rain-rate increases tend to be smaller. Close to the centre of a tropical cyclone, rain-rate increases can exceed the 7% per degree of warming indicated by the Clausius–Clapeyron scaling. In recent global simulations, the maximum increase in rain rate was estimated to be about 10% per degree of warming<sup>25</sup>. Therefore, anthropogenic warming is expected to increase the rain rates and decrease the translation speeds of tropical cyclones. Because the amount of local tropical-cyclone-related rainfall depends on both rain rate and translation speed (with a decrease in translation speed having about the same local effect, proportionally, as an increase in rain rate, as noted above), each of these two independent effects of anthropogenic warming is expected to increase local rainfall. The increase in local rainfall caused by a 10% global increase in the tropical-cyclone rain rate per degree of warming, as predicted in numerical simulations<sup>25</sup>, would be doubled by a concurrent slowdown in tropical-cyclone translation speed of as little as 10%. In addition, increases in tropical-cyclone rain rate due to warming become smaller further from the tropical-cyclone centre, such that rain-band rain rates increase by only a fraction of the simulated increase in rain-rates under the eye-wall<sup>10–12,25</sup>. There is no such relationship for translation speed, so a slowdown in translation speed will increase local rainfall amounts by the same percentage at all distances from the tropical-cyclone centre. This further strengthens the potential for the effect of past and projected changes in translation speed to dominate that of changes in rain rate.

<sup>1</sup>NOAA National Centers for Environmental Information, Center for Weather and Climate, Madison, WI, USA. e-mail: james.kossin@noaa.gov



**Fig. 1 | Global and hemispheric time series of annual-mean tropical-cyclone translation speed and their linear trends.** **a, b,** The period of the time series is 1949–2016. Gray shading indicates the two-sided 95% confidence bounds of the trends, which have been adjusted for autocorrelation as needed (see Methods). Time series are shown for the global data (**a**) and for the data separated by Northern and Southern hemispheres (**b**). The trends for the global and Northern Hemisphere data have confidence levels of about 100% (based on *P* values; see Extended Data Table 1). The trend in the Southern Hemisphere is not significant at the two-sided 95% confidence level.

Time series of annual-mean global and hemispheric translation speed are shown in Fig. 1, based on global tropical-cyclone ‘best-track’ data (see Methods). A highly significant global slowdown of tropical-cyclone translation speed is evident, of  $-10\%$  over the 68-yr period 1949–2016 (Extended Data Table 1). During this period, global-mean surface temperature has increased by about  $0.5^{\circ}\text{C}^9$ . The global distribution of translation speed exhibits a clear shift towards slower speeds in the second half of the 68-yr period, and the differences are highly significant throughout most of the distribution (Fig. 2). This slowing is found in both the Northern and Southern Hemispheres (Fig. 1), but

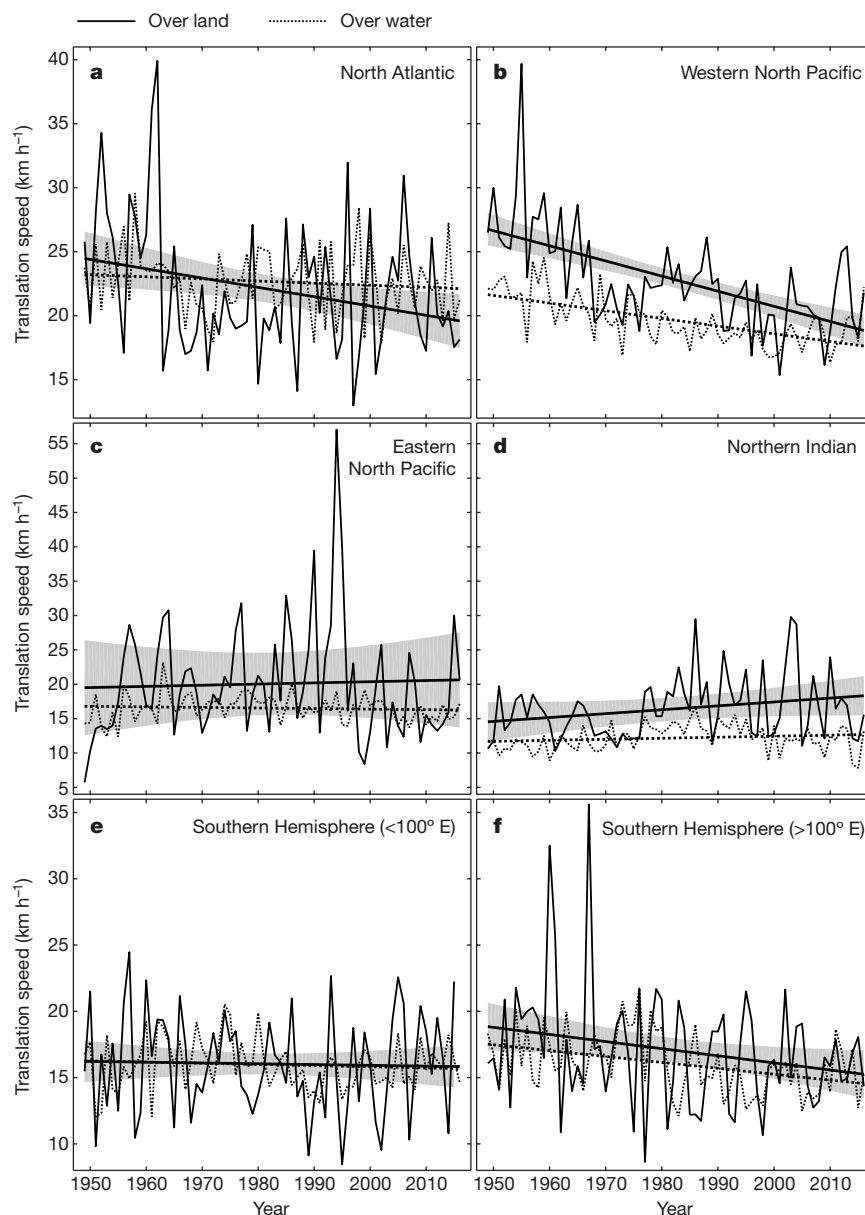


**Fig. 2 | Change in the global distribution of tropical-cyclone translation speed.** The changes are shown between the first and second halves of the period 1949–2016. Error bars show two-sided 95% confidence intervals based on bootstrap sampling (see Methods). There are significantly higher probabilities of translation speeds of less than  $20\text{ km h}^{-1}$  in the later period and significantly higher probabilities of translations speeds of greater than  $20\text{ km h}^{-1}$  in the earlier period.

is stronger and more significant in the Northern Hemisphere, where the annual number of tropical cyclones is generally greater. The time series for the Southern Hemisphere exhibits a change-point around 1980, but the reason for this is not clear. Before 1980, analyses of tropical cyclones depended largely on polar orbiting satellites to provide estimates of tropical-cyclone centre position; after 1980, geostationary satellites were also available. Such a change in the availability of satellite information is expected to introduce heterogeneity into estimates of tropical-cyclone intensity<sup>26</sup>, but estimates of tropical-cyclone position should be comparatively insensitive to such changes. In addition, estimates of translation speed should be comparatively insensitive to less-than-perfect estimates of tropical-cyclone position along the tropical-cyclone track because the speed is calculated between each pair of positions, and position errors along the track should mostly cancel each other when the speeds are averaged along the track.

The trends in tropical-cyclone translation speed and their signal-to-noise ratios vary considerably when the data are parsed by region, but slowing is found in every basin except the northern Indian Ocean (Extended Data Fig. 1, Extended Data Table 1). Significant slowings of  $-20\%$  in the western North Pacific Ocean and of  $-15\%$  in the region around Australia (Southern Hemisphere, east of  $100^{\circ}\text{E}$ ) are observed. When the data are constrained within global latitude belts (Extended Data Fig. 2, Extended Data Table 1), significant slowing is observed at latitudes above  $25^{\circ}\text{N}$  and between  $0^{\circ}$  and  $30^{\circ}\text{S}$ . Slowing trends near the equator tend to be smaller and not significant, whereas there is a substantial (but insignificant) increasing trend in translation speed at higher latitudes in the Southern Hemisphere.

When only that data that correspond to tropical cyclones over water are considered, which amounts to about 90% of the global best-track data, the trend statistics are indistinguishable from the global slowing trends (Extended Data Table 1). The 10% of the global data that correspond to tropical cyclones over land, where local rainfall effects become more societally relevant, also exhibit a slowing trend, but it is not significant. However, changes in tropical-cyclone translation speed over land vary substantially by region (Fig. 3, Extended Data Table 1). There is a substantial and significant slowing trend over land areas



**Fig. 3 | Time series of annual-mean tropical-cyclone translation speed and their linear trends over land and water. a–f,** Time series are shown for the individual ocean basins, over land (solid lines) and water (dotted lines). Grey shading indicates the two-sided 95% confidence bounds of the trends over land, corrected for autocorrelation as needed. The regions

are the North Atlantic (a), western North Pacific (b), eastern North Pacific (c) and northern Indian (d) basins, and the regions west (e) and east (f) of 100° E in the Southern Hemisphere. The regions over land that are affected in each region are shown in Extended Data Fig. 3.

affected by North Atlantic tropical cyclones (20% reduction over the 68-yr period), by western North Pacific tropical cyclones (30% reduction) and by tropical cyclones in the Australian region (19% reduction, but the significance is marginal). These trends have almost certainly increased local rainfall totals in these regions, which are more difficult to measure directly. Contrarily, the tropical-cyclone translation speeds over land areas affected by eastern North Pacific and northern Indian tropical cyclones, and of tropical cyclones that have affected Madagascar and the east coast of Africa, all exhibit positive trends, although none is significant. Note that several of the time series shown here exhibit outliers that suggest non-normality of the residuals from the trend lines; a discussion and analysis of the robustness of the trends presented here is provided in Methods.

None of the analyses presented here has any dependence on tropical-cyclone intensity; in some cases, the translation speeds were calculated in the absence of any concurrent intensity estimates in the best-track data (cases where location but not intensity information is available).

This allows much greater confidence in the homogeneity of the historical best-track data, but does not allow for intensity-based stratifications of the data. This is a caveat of these analyses because tropical-cyclone rain rates have been shown to be a function of their intensity, with greater rates linked to stronger tropical cyclones<sup>27</sup>. In this case, there is a possibility for offsetting the effects of tropical-cyclone slowdown if the trends in translation speed are dominated by weaker systems and for compounding these effects if the trends are dominated by stronger systems. This is left as an open question.

The analyses presented here do not constitute a detection and attribution study because there are likely to be many factors, natural and anthropogenic, that control tropical-cyclone translation speed. For example, the best-track data exhibit a global 10% reduction in translation speed during a period in which global-mean surface temperatures increased by about 0.5 °C; however, this finding does not provide a true measure of the climate sensitivity of these related phenomena. To determine the true sensitivity (that is, the expected change in translation

speed as a function of anthropogenic forcing), further analyses and numerical simulations are required.

In addition to the global slowing of tropical-cyclone translation speed identified here, there is evidence that tropical cyclones have migrated poleward in several regions<sup>28</sup>. Of particular relevance here, the rate of migration in the western North Pacific was found to be large, which has had a substantial effect on regional tropical-cyclone-related hazard exposure<sup>29</sup>. When this finding is considered in tandem with the substantial slowdown of translation speed over land in this region (30% since 1949), the potential for increased hazard exposure becomes greater still, particularly to fresh-water flooding hazards, which can pose an especially large mortality risk<sup>30</sup>. Further compounding these changes in regional exposure, the projected increases in tropical-cyclone rain rate in the western North Pacific for the late twenty-first century are about twice the projected global-mean increase<sup>25</sup>. These recently identified trends in tropical-cyclone track behaviour emphasize that tropical-cyclone frequency and intensity should not be the only metrics considered when establishing connections between climate variability and change and the risks associated with tropical cyclones, both past and future. These trends further support the idea that the behaviours of tropical cyclones are being altered in societally relevant ways by anthropogenic factors. Continued research into the connections between tropical cyclones and climate is essential to understanding and predicting the changes in risk that are occurring on a global scale.

The analyses presented here demonstrate changes in the behaviour of translation speed, but local rainfall totals are also affected by translation direction. For example, a tropical cyclone that follows a looping track over some region could be translating quickly along the loop, but the rainfall totals in the region would still be large owing to the spatially confined track. In 2017, Hurricane Harvey not only translated slowly over Texas but also reversed direction and thus affected the same region over a particularly long duration. There is currently no formal definition of what constitutes a 'stalled track', although this term has been used to describe the track of Hurricane Harvey. Future studies that focus on tropical cyclones that remain geographically constrained for extended durations are warranted.

## Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0158-3>.

Received: 22 February 2018; Accepted: 18 April 2018;  
Published online 6 June 2018.

- Mann, M. E. et al. Influence of anthropogenic climate change on planetary wave resonance and extreme weather events. *Sci. Rep.* **7**, 19831 (2017).
- Held, I. M. & Soden, B. J. Robust responses of the hydrological cycle to global warming. *J. Clim.* **19**, 5686–5699 (2006).
- He, J. & Soden, B. J. Anthropogenic weakening of the tropical circulation: the relative roles of direct CO<sub>2</sub> forcing and sea surface temperature change. *J. Clim.* **28**, 8728–8742 (2015).
- Coumou, D., Lehmann, J. & Beckmann, J. The weakening summer circulation in the Northern Hemisphere mid-latitudes. *Science* **348**, 324–327 (2015).
- Vecchi, G. A. & Soden, B. J. Global warming and the weakening of the tropical circulation. *J. Clim.* **20**, 4316–4340 (2007).
- Vecchi, G. A. et al. Weakening of tropical Pacific atmospheric circulation due to anthropogenic forcing. *Nature* **441**, 73–76 (2006).
- Grise, K. M. & Polvani, L. M. Understanding the timescales of the tropospheric circulation response to abrupt CO<sub>2</sub> forcing in the Southern Hemisphere: seasonality and the role of the stratosphere. *J. Clim.* **30**, 8497–8515 (2017).

- He, C., Wu, B., Zou, L. & Zhou, T. Responses of the summertime Subtropical Anticyclones to global warming. *J. Clim.* **30**, 6465–6479 (2017).
- Hartmann, D. L. et al. in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Stocker, T. F. et al.) 159–254 (Cambridge Univ. Press, Cambridge, 2013).
- Christensen, J. H. et al. in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Stocker, T. F. et al.) 1217–1308 (Cambridge Univ. Press, Cambridge, 2013).
- Kossin, J. P. et al. in *Climate Science Special Report: Fourth National Climate Assessment Vol. I* (eds Wuebbles, D. J. et al.) 257–276 (US Global Change Research Program, Washington DC, 2017).
- Walsh, K. J. E. et al. Tropical cyclones and climate change. *Wiley Interdiscip. Rev. Clim. Change* **7**, 65–89 (2016).
- Emanuel, K. A. Assessing the present and future probability of Hurricane Harvey's rainfall. *Proc. Natl Acad. Sci. USA* **114**, 12681–12684 (2017).
- Risser, M. D., & Wehner, M. F. Attributable human-induced changes in the likelihood and magnitude of the observed extreme precipitation during Hurricane Harvey. *Geophys. Res. Lett.* **44**, 12457–12464 (2017).
- van Oldenborgh, G. J. et al. Attribution of extreme rainfall from Hurricane Harvey, August 2017. *Environ. Res. Lett.* **12**, 124009 (2017); corrigendum **13**, 019501 (2018).
- DeMaria, M., Mainelli, M., Shay, L. K., Knaff, J. A. & Kaplan, J. Further improvement to the Statistical Hurricane Intensity Prediction Scheme (SHIPS). *Weather Forecast.* **20**, 531–543 (2005).
- Kossin, J. P. & Sitkowski, M. Predicting hurricane intensity and structure changes associated with eyewall replacement cycles. *Weather Forecast.* **27**, 484–488 (2012).
- Hegerl, G. E. et al. Challenges in quantifying changes in the global water cycle. *Bull. Am. Meteorol. Soc.* **96**, 1097–1115 (2015).
- Allen, M. R. & Ingram, W. J. Constraints on future changes in climate and the hydrologic cycle. *Nature* **419**, 224–232 (2002).
- Pendergrass, A. G. & Hartmann, D. L. The atmospheric energy constraint on global-mean precipitation change. *J. Clim.* **27**, 757–768 (2014).
- Allan, R. F. & Soden, B. J. Atmospheric warming and the amplification of precipitation extremes. *Science* **321**, 1481–1484 (2008).
- Kitoh, A. & Endo, H. Changes in precipitation extremes projected by a 20-km mesh global atmospheric model. *Weather Clim. Extrem.* **11**, 41–52 (2016).
- Kharin, V. V., Zwiers, F. W., Zhang, X. & Wehner, M. Changes in the temperature and precipitation extremes in the CMIP5 ensemble. *Clim. Change* **119**, 345–357 (2013).
- Prein, A. F. et al. The future intensification of hourly precipitation extremes. *Nat. Clim. Chang.* **7**, 48–52 (2017).
- Knutson, T. R. et al. Global projections of intense tropical cyclone activity for the late twenty-first century from dynamical downscaling of CMIP5/RCP4.5 scenarios. *J. Clim.* **28**, 7203–7224 (2015).
- Kossin, J. P., Olander, T. L. & Knapp, K. R. Trend analysis with a new global record of tropical cyclone intensity. *J. Clim.* **26**, 9960–9976 (2013).
- Lonfat, M., Marks, F. D. & Chen, S. Precipitation distribution in tropical cyclones using the Tropical Rainfall Measuring Mission (TRMM) microwave imager: a global perspective. *Mon. Weath. Rev.* **132**, 1645–1660 (2004).
- Kossin, J. P., Emanuel, K. A. & Vecchi, G. A. The poleward migration of the location of tropical cyclone maximum intensity. *Nature* **509**, 349–352 (2014).
- Kossin, J. P., Emanuel, K. A. & Camargo, S. J. Past and projected changes in western North Pacific tropical cyclone exposure. *J. Clim.* **29**, 5725–5739 (2016).
- Rappaport, E. N. Loss of life in the United States associated with recent Atlantic tropical cyclones. *Bull. Am. Meteorol. Soc.* **81**, 2065–2073 (2000).

**Acknowledgements** This work was supported by NOAA's National Centers for Environmental Information.

**Reviewer information** Nature thanks M. Mann and C. Patricola for their contribution to the peer review of this work.

**Competing interests** The author declares no competing interests.

## Additional information

**Extended data** is available for this paper at <https://doi.org/10.1038/s41586-018-0158-3>.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>.

**Correspondence and requests for materials** should be addressed to J.P.K.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## METHODS

Best-track data are taken from IBTrACS<sup>31</sup> (see Methods section 'Data availability'). On the basis of comparisons of IBTrACS data sources<sup>32</sup>, data from the US National Hurricane Center (NHC) and Joint Typhoon Warning Center (JTWC) were combined to provide global coverage. NHC data cover the North Atlantic and eastern North Pacific oceans. JTWC data cover the western North Pacific and northern Indian oceans and the Southern Hemisphere, which includes the southern Indian and South Pacific oceans. 100° E was used to separate tropical cyclones that affect the Australia region from those that affect Madagascar and the east coast of Africa. The analyses presented here are not highly sensitive to this choice. The period 1949–2016 was chosen on the basis of uniform availability of data from each region.

Time series are based on annual-mean translation speeds. Translation speed is calculated using neighbouring positions along each tropical-cyclone track (these are provided in six-hourly intervals throughout the lifetime of each tropical cyclone). Distances between locations are calculated along a great circle arc. Trends are estimated by linear regression. The *P* values of the regression are based on the full degrees of freedom in the 68-year time series. Because some of the time series exhibit autoregressive (AR(1)) persistence, as determined from a Durbin–Watson test, confidence intervals are provided with degrees of freedom adjusted when needed. Statistical significance is based on the two-sided 95% confidence intervals (not the *P* values).

The percentage change is calculated by dividing the difference between the last and first points of the best-fit trend line by the first point. Over-land positions are determined using a high-resolution global topography map (see Methods section 'Data availability'). If either one or both of the two locations used to calculate the translation speed are over land, then the speed is considered to be an over-land speed. The two-sided 95% confidence bounds on the probability-density histogram (Fig. 2) are constructed by repeated random sampling with replacement within each of the two time periods. The error bars show  $\pm 2\sigma$  from the mean in each histogram bin.

The final post-season-reanalysed NHC and JTWC data for 2017 were not yet available at the time of writing this manuscript, but it is interesting to identify Hurricane Harvey's effect on the 2017-mean tropical-cyclone translation speed over land that is affected by Atlantic hurricanes. Using the 'operational best-track' data from the Automated Tropical Cyclone Forecasting System (ATCF)<sup>33</sup>, the 2017-mean over-land Atlantic translation speed is 17.9 km h<sup>-1</sup>, which is at the slowest 20th percentile of over-land translation speeds for the period since 1949. Adding this point to the annual time series slightly increases the magnitude of the slowing trend and decreases the *P* value of the regression to 0.0200 (from 0.0275).

To test the robustness of the trends shown here, particularly to outliers, which are evident in several of the time series shown and could indicate a non-Gaussian distribution of the regression residuals, all trends were recalculated using the  $L^1$  norm in place of the  $L^2$  (ordinary least-squares) norm (Extended Data Table 2). A few of the trends are slightly affected by this choice, but none of the significant trends from the  $L^2$  regressions becomes insignificant. Still, the trend statistics pre-

sented here should be interpreted with the understanding that the distribution of the regression residuals may deviate from normal (Extended Data Figs. 4, 5), particularly when the global data are parsed to smaller subregions. Caution in interpretation is particularly important for the trends in over-land translation speed in the Australian region, which exhibit two clear outliers in the early part of the time series and for which the significance is marginal. On the other hand, the slowing trend in the Australian region as a whole (Extended Data Table 1) is highly significant ( $P=0.005$ ) and of similar magnitude to the over-land trend, which provides additional confidence that the trend found in a small subsample of those data is physical and not random. Similarly, the global trend (Fig. 1a) is not very sensitive to the removal of individual points or to small changes in the start and end points of the time series, but these sensitivities increase, as expected, when the global data are subsampled down to progressively finer regional scales.

The poleward migration of tropical cyclones discussed above could be related to the slowdown of translation speed via tropical-cyclone 'beta drift'<sup>34</sup>. In particular, a poleward migration would reduce beta drift, which, all other things being equal, would reduce the translation speed. In addition, a decrease in mean tropical-cyclone intensity would reduce the translation speed through beta drift; however, no such decrease has been observed in the period considered here and there is some evidence that mean tropical-cyclone intensity has increased since the early 1980s<sup>26,35</sup>. The potential relationships, based on beta-drift arguments, between the translation speed, poleward migration and intensity of tropical cyclones warrants further study.

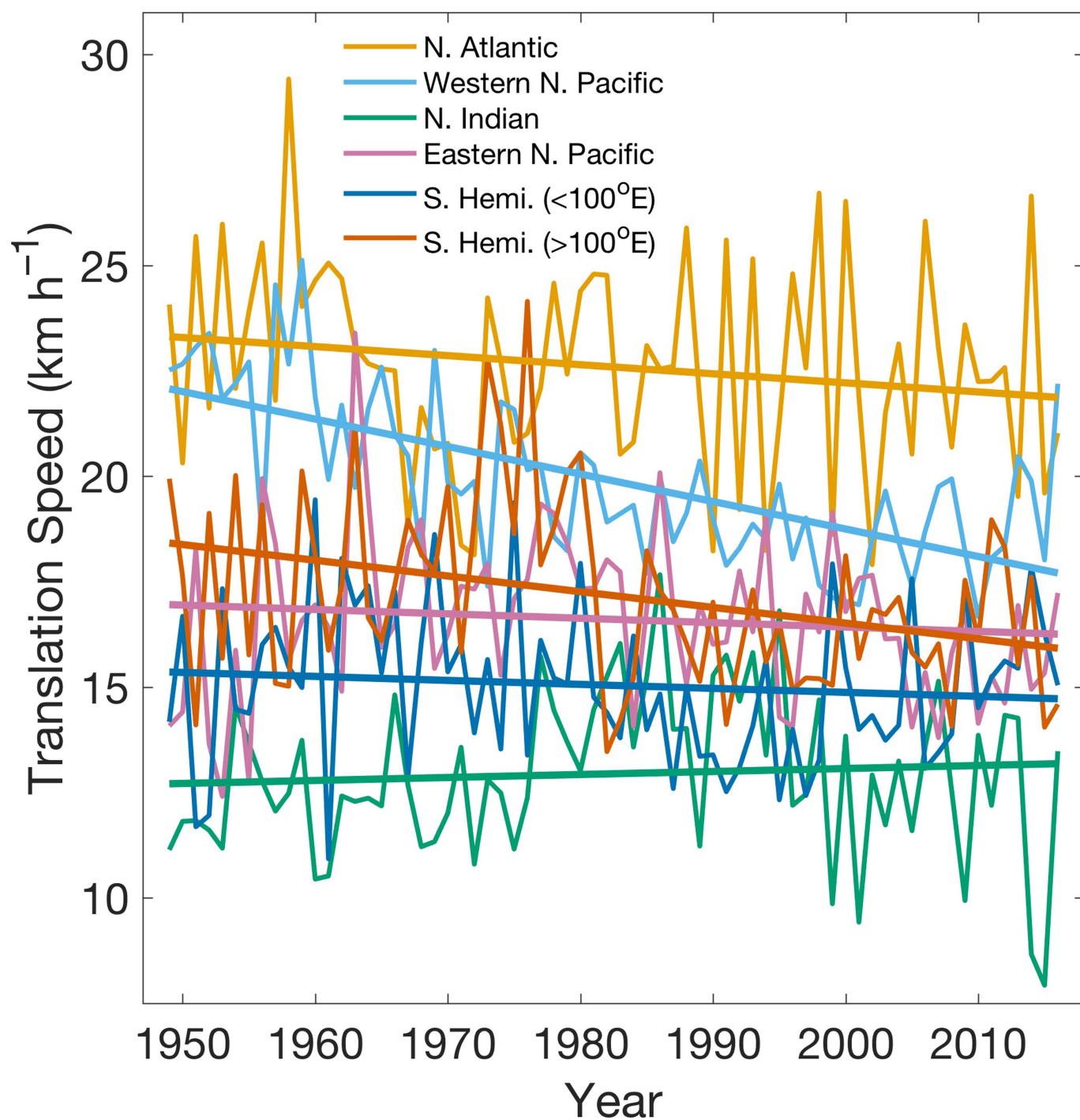
The colour choices in Extended Data Figs. 2 and 3 follow a specification designed to mitigate colour blindness<sup>36</sup>.

No statistical methods were used to predetermine sample size.

**Data availability.** The tropical-cyclone data analysed during this study were taken from the International Best Track Archive for Climate Stewardship (IBTrACS; <https://www.ncdc.noaa.gov/ibtracs/>, file 'Allstorms.ibtracs\_all.v03r10.nc'). The over-land tropical-cyclone positions were determined from the 2-Minute Gridded Global Relief Data (ETOPO2v2; <https://www.ngdc.noaa.gov/mgg/global/etopo2.html>). These data are also available from the corresponding author on request.

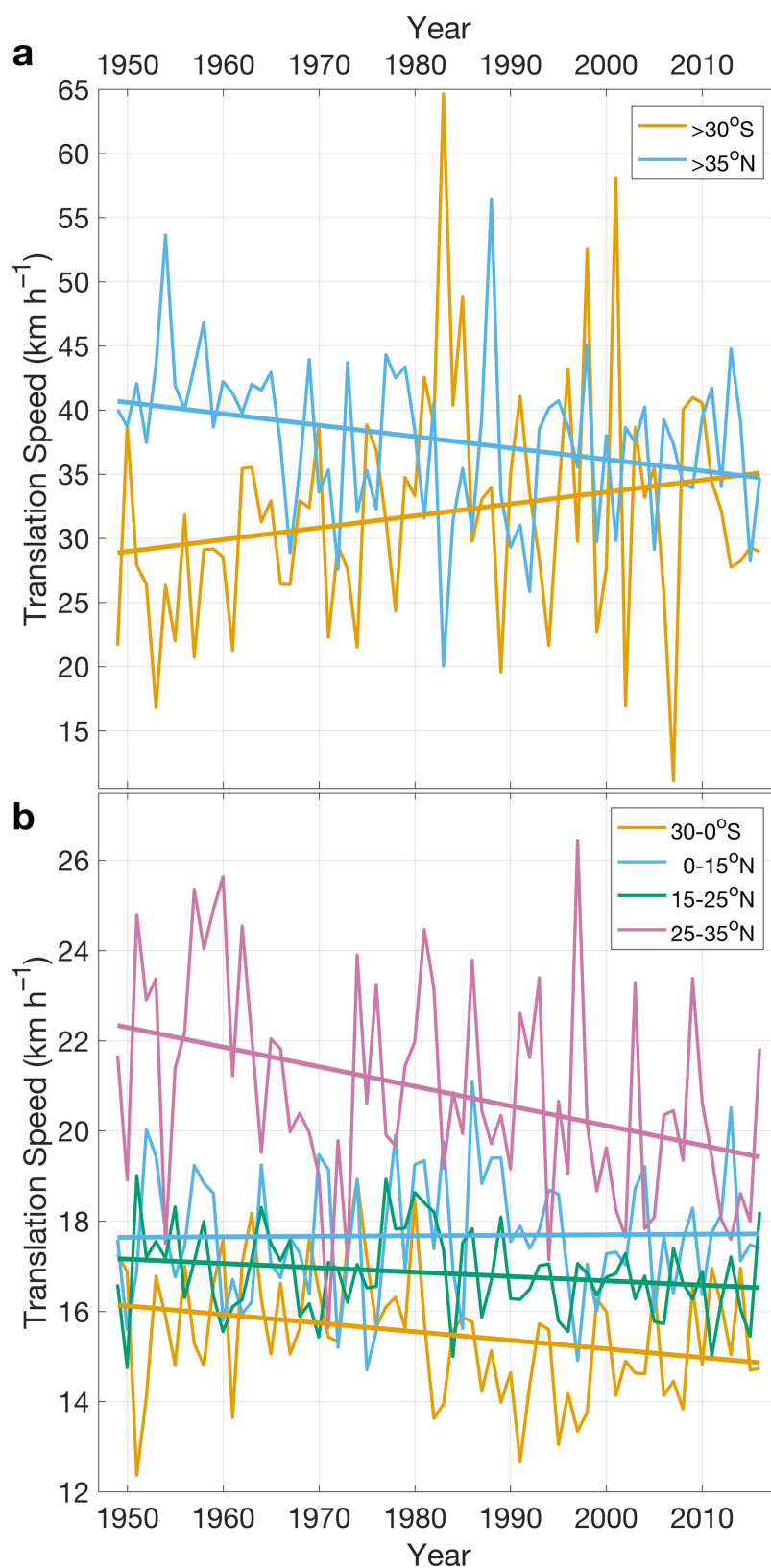
**Code availability.** All codes used to read, analyse and plot the data are available from the corresponding author on request.

31. Knapp, K. R., Kruk, M. C., Levinson, D. H., Diamond, H. J. & Neumann, C. J. The International Best Track Archive for Climate Stewardship (IBTrACS). *Bull. Am. Meteorol. Soc.* **91**, 363–376 (2010).
32. Schreck, C. J. III, Knapp, K. R. & Kossin, J. P. The impact of best track discrepancies on global tropical cyclone climatologies using IBTrACS. *Mon. Weath. Rev.* **142**, 3881–3899 (2014).
33. Sampson, C. R. & Schrader, A. J. The automated tropical cyclone forecasting system (version 3.2)". *Bull. Am. Meteorol. Soc.* **81**, 1231–1240 (2000).
34. Smith, R. B. A hurricane beta-drift law. *J. Atmos. Sci.* **50**, 3213–3215 (1993).
35. Elsner, J. B., Kossin, J. P. & Jagger, T. H. The increasing intensity of the strongest tropical cyclones. *Nature* **455**, 92–95 (2008).
36. Wong, B. Points of view: Color blindness. *Nat. Methods* **8**, 441 (2011).



**Extended Data Fig. 1 | Time series of annual-mean tropical-cyclone translation speed and their linear trends in varying ocean basins.** Time series are shown for the individual ocean basins: North Atlantic, western

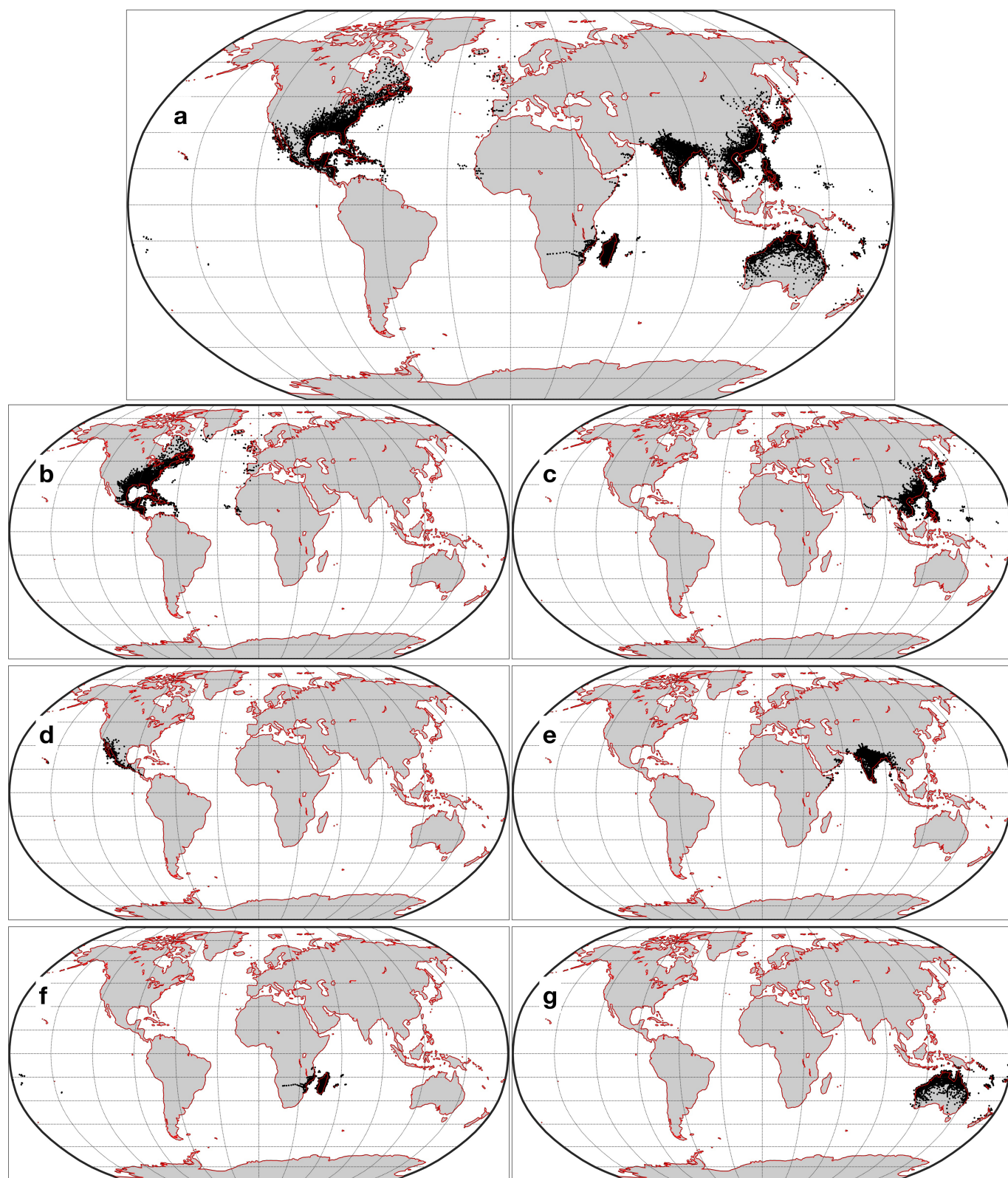
North Pacific, northern Indian, eastern North Pacific, and the regions west and east of  $100^\circ\text{E}$  in the Southern Hemisphere.



**Extended Data Fig. 2 | Time series of annual-mean tropical-cyclone translation speed and their linear trends in varying latitude belts.**

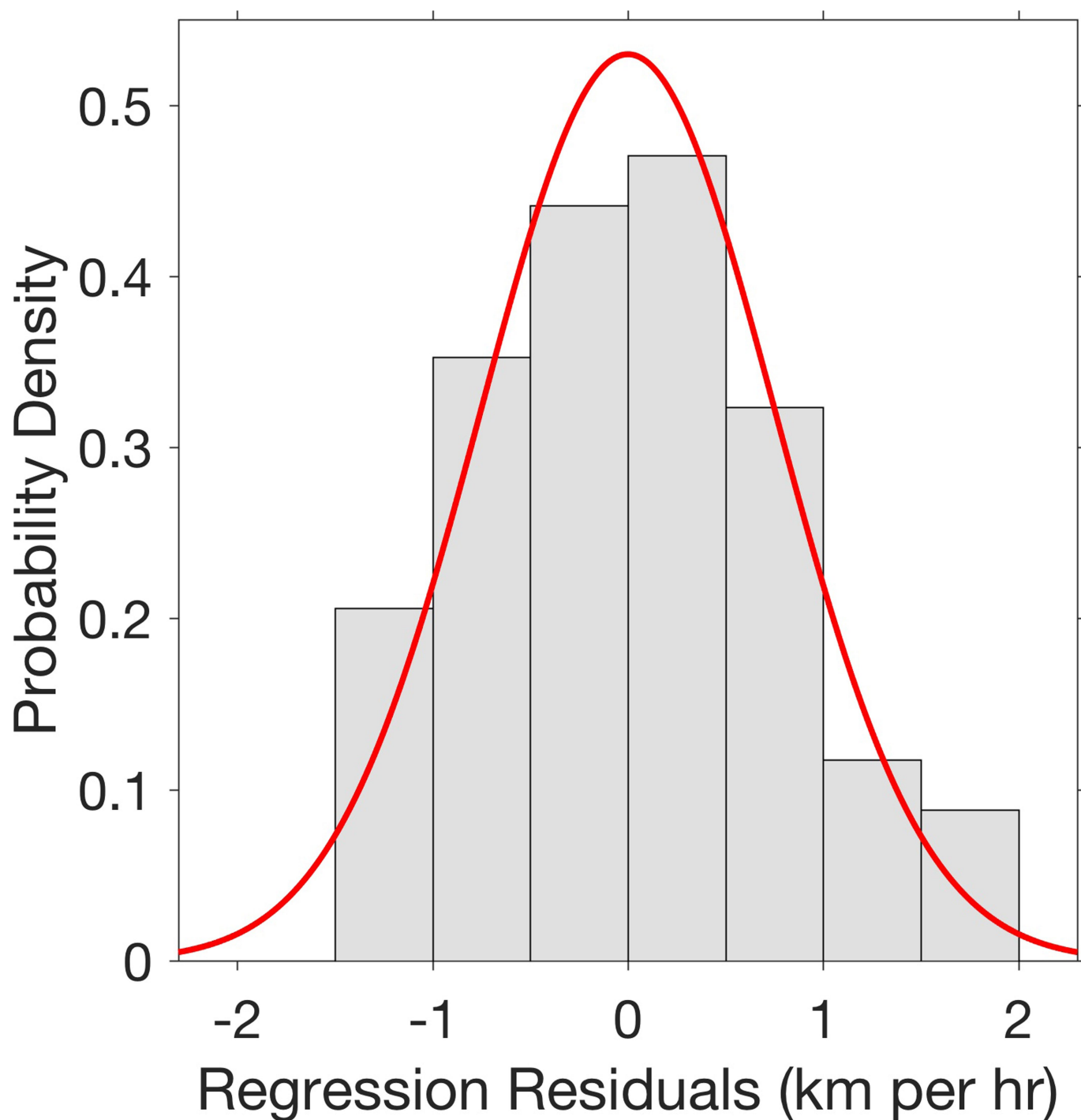
**a, b,** Time series are shown for global latitude belts southward (poleward) of  $30^{\circ}\text{S}$  and northward (poleward) of  $35^{\circ}\text{N}$  (**a**), and in the belts  $30^{\circ}\text{--}0^{\circ}\text{S}$ ,

$0^{\circ}\text{--}15^{\circ}\text{N}$ ,  $15^{\circ}\text{--}25^{\circ}\text{N}$  and  $25^{\circ}\text{--}35^{\circ}\text{N}$  (**b**). The broader latitude belts defined in the less-active Southern Hemisphere were needed to have a large enough sample of tropical cyclones in each to perform the analyses. The analyses shown here are fairly robust to the choice of latitude-belt bounds.



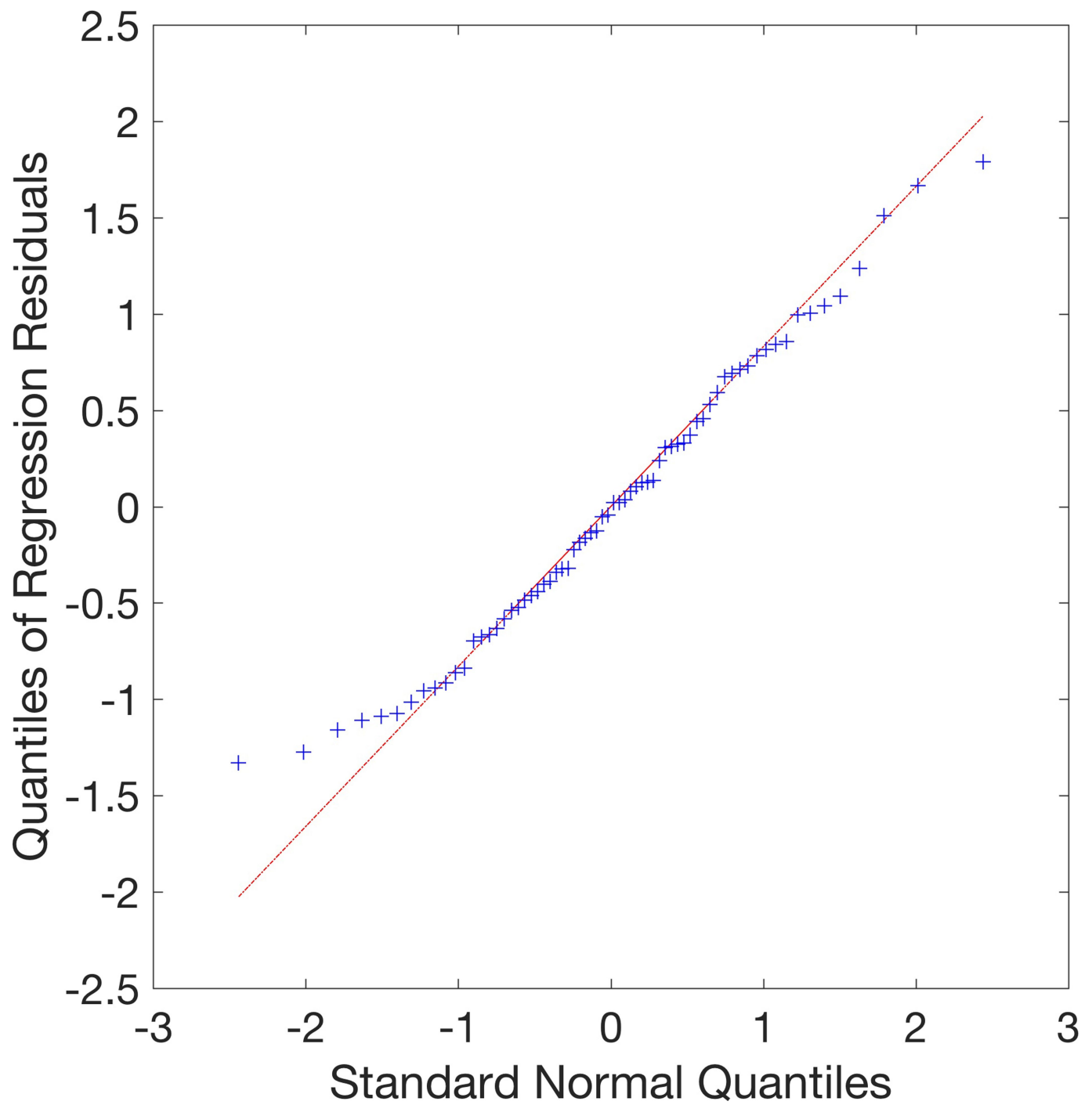
**Extended Data Fig. 3 | Best-track tropical-cyclone centre locations for tropical-cyclone translation speeds over land.** Locations (black dots) are shown for the regions considered in Fig. 3 and Extended Data Table 1. **a–g,** Over-land positions are shown for land areas affected by tropical

cyclones globally (**a**), in the North Atlantic (**b**), western North Pacific (**c**), eastern North Pacific (**d**) and northern Indian (**e**) basins, and in the regions west (**f**) and east (**g**) of 100° E in the Southern Hemisphere.



**Extended Data Fig. 4 | Distribution of the residuals from the ordinary least-squares regression.** The histogram shows the distribution of residuals from the trend for the time series of annual-mean global tropical-

cyclone translation speed shown in Fig. 1a. The normal distribution is shown in red.



**Extended Data Fig. 5 | Quantile–quantile plot for the residuals of the global ordinary least-squares regression.** The left tail of the distribution is thinner than the normal distribution, which suggests that the likelihood

of an extremely slow translation speed is less than would be found in a normally distributed sample.

Extended Data Table 1 | Trends in tropical-cyclone translation speed and their statistics

	Trend (km h <sup>-1</sup> yr <sup>-1</sup> )	Change (%)	p-value (uncorrected)	95% confidence interval of the trend
Global	<b>-0.03</b>	<b>-10</b>	<10 <sup>-7</sup>	[-0.04, -0.02]*
N. Hemisphere	<b>-0.03</b>	<b>-11</b>	<10 <sup>-6</sup>	[-0.05, -0.02]
S. Hemisphere	-0.02	-9	0.02	[-0.05, 0.002]*
N. Atlantic	-0.02	-6	0.16	[-0.05, 0.008]
Western N. Pacific	<b>-0.07</b>	<b>-20</b>	<10 <sup>-8</sup>	[-0.08, -0.05]
Eastern N. Pacific	-0.01	-4	0.39	[-0.03, 0.01]
N. Indian	+0.01	+4	0.55	[-0.02, 0.04]*
S. Hemi. (<100°E)	-0.01	-4	0.41	[-0.03, 0.01]
S. Hemi. (≥100°E)	<b>-0.04</b>	<b>-15</b>	0.005	[-0.06, -0.01]
>35°N	<b>-0.09</b>	<b>-15</b>	0.02	[-0.16, -0.02]
25–35°N	<b>-0.04</b>	<b>-13</b>	0.0025	[-0.07, -0.02]
15–25°N	-0.01	-4	0.10	[-0.02, 0.002]
0–15°N	0.0	0	0.90	[-0.02, 0.02]
0–30°S	<b>-0.02</b>	<b>-8</b>	0.03	[-0.04, -0.003]
> 30°S	+0.10	+23	0.09	[-0.01, 0.21]
Global (water)	<b>-0.03</b>	<b>-11</b>	<10 <sup>-7</sup>	[-0.04, -0.02]*
Global (land)	-0.01	-4	0.41	[-0.05, 0.03]*
N. Atlantic (land)	<b>-0.07</b>	<b>-20</b>	0.03	[-0.14, -0.01]
W. N. Pacific (land)	<b>-0.12</b>	<b>-30</b>	<10 <sup>-7</sup>	[-0.16, -0.08]
E. N. Pacific (land)	+0.02	+6	0.75	[-0.14, 0.17]*
N. Indian (land)	+0.06	+26	0.04	[-0.01, 0.12]*
S. Hemi. (<100°E, land)	+0.02	+7	0.53	[-0.04, 0.07]
Madagascar				
S. Hemi. (≥100°E, land)	<b>-0.05</b>	<b>-19</b>	0.05	[-0.11, 0.00]
Australia				

The changes are shown as trends (km h<sup>-1</sup> yr<sup>-1</sup>) and in percentage change over the period 1949–2016 in various regions and latitude belts. The regions include the North Atlantic, the western and eastern North Pacific and the northern Indian Ocean basins, and the Southern Hemisphere separated by 100° E. Other regions are separated by data over land and water, and by global latitude belts. Significance is shown by *P* values of the regressions and confidence intervals of the trends. Where identified by an asterisk, confidence intervals are based on effective degrees of freedom adjusted for serial correlation (see Methods). Significant trends, determined on the basis of the confidence interval, after correcting for serial correlation as needed, are shown in bold.

Extended Data Table 2 | Trends in tropical-cyclone translation speed and their statistics under the  $L^1$  norm

	Trend (km h <sup>-1</sup> yr <sup>-1</sup> )	Change (%)	95% confidence interval of the trend
Global	<b>-0.03</b>	<b>-9</b>	[-0.04, -0.02]*
N. Hemisphere	<b>-0.03</b>	<b>-10</b>	[-0.04, -0.02]
S. Hemisphere	<b>-0.04</b>	<b>-14</b>	[-0.06, -0.01]*
N. Atlantic	-0.01	-4	[-0.04, 0.02]
Western N. Pacific	<b>-0.08</b>	<b>-22</b>	[-0.09, -0.06]
Eastern N. Pacific	-0.02	-7	[-0.04, 0.006]
N. Indian	<b>+0.03</b>	<b>+18</b>	[0.00, 0.06]*
S. Hemi. (<100°E)	-0.02	-9	[-0.04, 0.003]
S. Hemi. (≥100°E)	<b>-0.05</b>	<b>-17</b>	[-0.08, -0.02]*
>35°N	-0.06	-9	[-0.13, 0.02]
25–35°N	<b>-0.06</b>	<b>-17</b>	[-0.08, -0.03]
15–25°N	-0.01	-4	[-0.02, 0.001]
0–15°N	+0.00	+0	[-0.01, 0.02]
0–30°S	<b>-0.03</b>	<b>-11</b>	[-0.05, -0.008]*
> 30°S	+0.10	+24	[-0.01, 0.21]
Global (water)	<b>-0.03</b>	<b>-10</b>	[-0.04, -0.02]*
Global (land)	-0.00	-1	[-0.04, 0.04]*
N. Atlantic (land)	<b>-0.06</b>	<b>-19</b>	[-0.13, -0.001]
W. N. Pacific (land)	<b>-0.10</b>	<b>-27</b>	[-0.15, -0.06]*
E. N. Pacific (land)	-0.00	-1	[-0.16, 0.15]*
N. Indian (land)	+0.03	+14	[-0.04, 0.10]*
S. Hemi. (<100°E, land)	-0.01	-3	[-0.06, 0.04]
Madagascar			
S. Hemi. (≥100°E, land)	<b>-0.05</b>	<b>-18</b>	[-0.10, 0.00]
Australia			

Similar to Extended Data Table 1, but using an  $L^1$  norm to calculate the trend statistics and test for robustness.

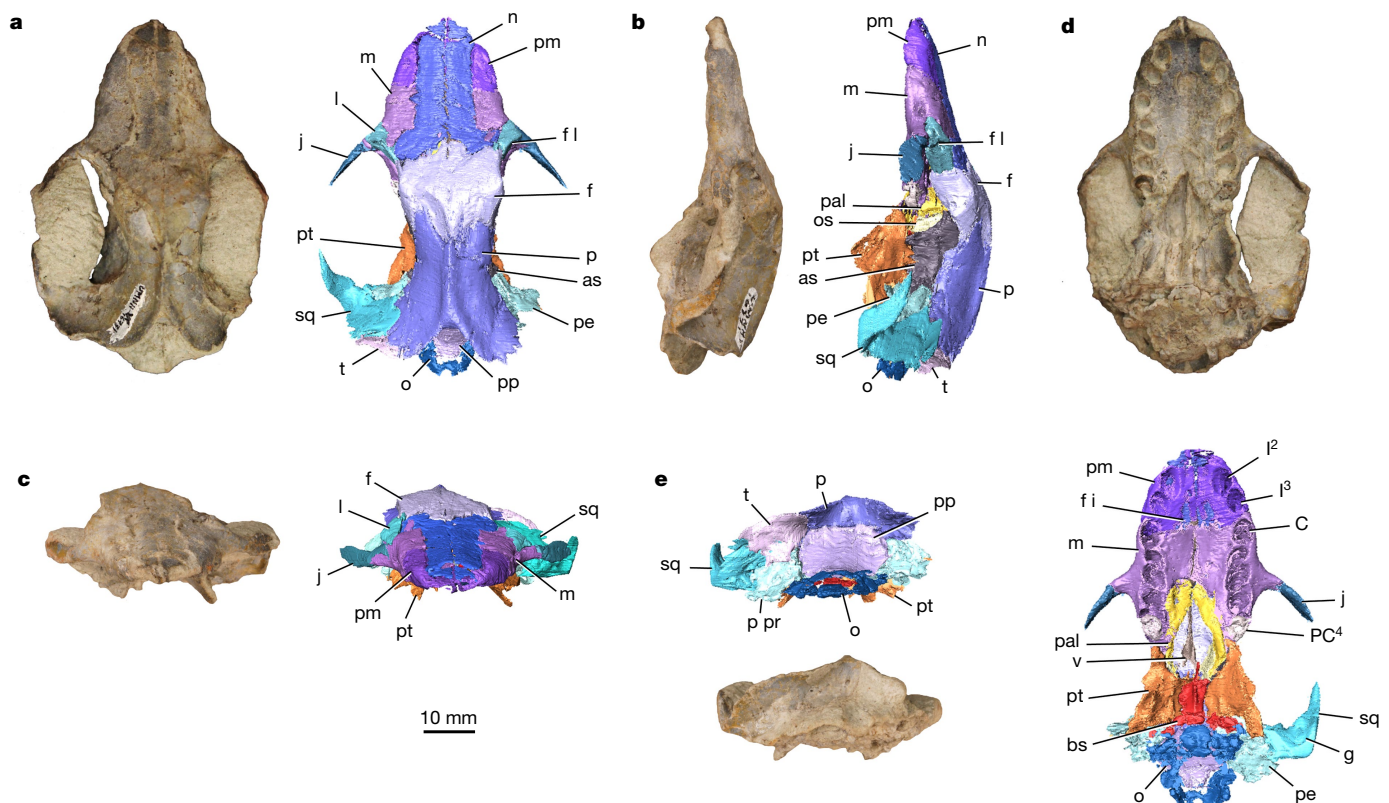
# Late-surviving stem mammal links the lowermost Cretaceous of North America and Gondwana

Adam K. Huttenlocker<sup>1\*</sup>, David M. Grossnickle<sup>2</sup>, James I. Kirkland<sup>3,4</sup>, Julia A. Schultz<sup>5</sup> & Zhe-Xi Luo<sup>2,5\*</sup>

Haramiyida was a successful clade of mammaliaforms, spanning the Late Triassic period to at least the Late Jurassic period, but their fossils are scant outside Eurasia and Cretaceous records are controversial<sup>1–4</sup>. Here we report, to our knowledge, the first cranium of a large haramiyidan from the basal Cretaceous of North America. This cranium possesses an amalgam of stem mammaliaform plesiomorphies and crown mammalian apomorphies. Moreover, it shows dental traits that are diagnostic of isolated teeth of supposed multituberculate affinities from the Cretaceous of Morocco, which have been assigned to the enigmatic ‘Hahnodontidae’. Exceptional preservation of this specimen also provides insights into the evolution of the ancestral mammalian brain. We demonstrate the haramiyidan affinities of Gondwanan hahnodontid teeth, removing them from multituberculates, and suggest that hahnodontid

mammaliaforms had a much wider, possibly Pangaeian distribution during the Jurassic–Cretaceous transition.

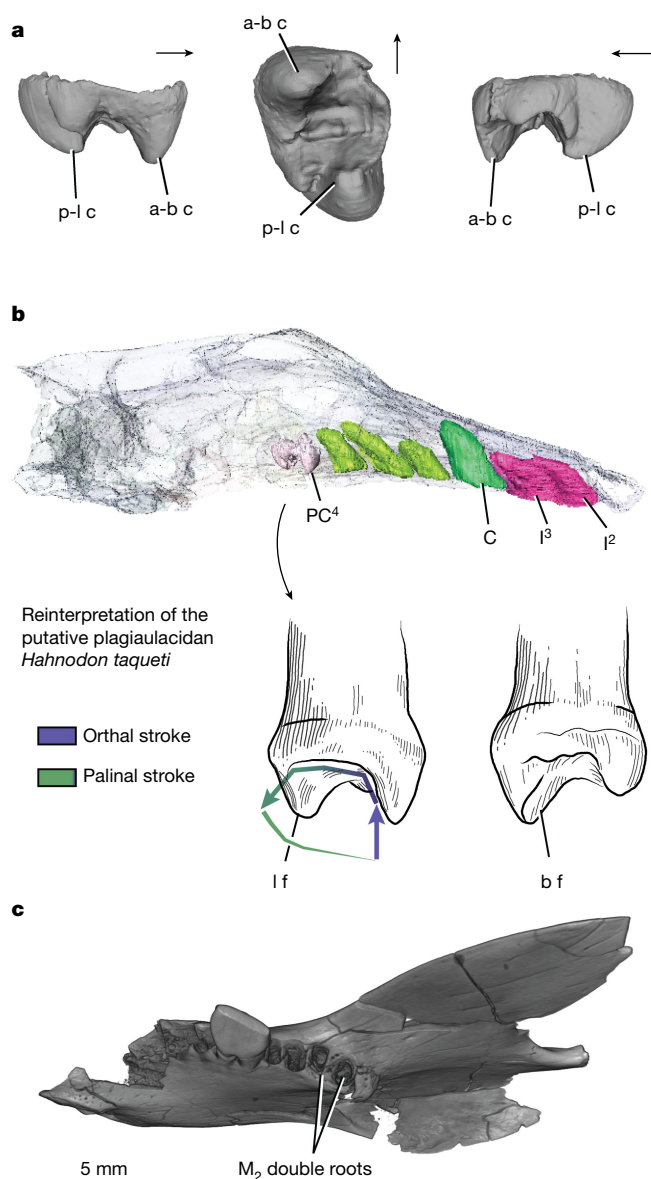
The ecological expansion of crown Mammalia accelerated in the wake of extinctions of diverse and disparate archaic mammalia-morph groups during the mid-Mesozoic<sup>1</sup>. Haramiyidans represent one such clade of early mammaliaforms that show preservation of notable links between nonmammalian and mammalian structure and physiology<sup>2–4</sup>. Their fossils, which were dated to the Late Triassic–Jurassic, were historically limited to isolated teeth and incomplete gnathal remains<sup>4,5</sup>, hindering a more complete understanding of their radiation. Recent discoveries of articulated skeletal material of eutherodont haramiyidans from the Middle–Upper Jurassic of China have shed new light on their diversification<sup>6,7</sup>, although these fossils have also sparked new debate over their phylogenetic position<sup>4,6–12</sup>.



**Fig. 1 | Cranium of *C. wahkarmoosuch*.** a–e, Holotype in dorsal (a), left lateral (b), frontal (c), ventral (d) and occipital (e) views. as, alisphenoid (epipterygoid); bs, basisphenoid; C, upper canine alveolus; f, frontal; f i, incisive foramen; f l, lacrimal foramen; g, squamosal glenoid; I<sup>2</sup>, first upper incisor alveolus (homologous to second incisor position in earlier

haramiyidans); I<sup>3</sup>, second upper incisor alveolus (homologous to third incisor position); j, jugal; l, lacrimal; m, maxilla; n, nasal; o, occipital; os, orbitosphenoid; p, parietal; p pr, paroccipital process; pal, palatine; PC<sup>4</sup>, in situ posterior upper postcanine (molar); pe, petrosal; pm, premaxilla; pp, postparietal; pt, pterygoid; sq, squamosal; t, tabular; v, vomer.

<sup>1</sup>Department of Integrative Anatomical Sciences, University of Southern California, Los Angeles, CA, USA. <sup>2</sup>Committee on Evolutionary Biology, The University of Chicago, Chicago, IL, USA. <sup>3</sup>Utah Geological Survey, Salt Lake City, UT, USA. <sup>4</sup>Natural History Museum of Utah, Salt Lake City, UT, USA. <sup>5</sup>Department of Organismal Biology and Anatomy, The University of Chicago, Chicago, IL, USA. \*e-mail: [ahuttenlocker@gmail.com](mailto:ahuttenlocker@gmail.com); [zxluo@uchicago.edu](mailto:zxluo@uchicago.edu)



**Fig. 2 | Dentition of *Cifelliodon*.** **a**, Right posterior molar in buccal (left), occlusal (middle) and lingual (right) views. Arrows denote anterior. **b**, Reinterpretation of *Hahnodon*. Roots and posterior molar of *Cifelliodon* (top) shown for comparison. Note that *Hahnodon* in **b** does not match completely divided  $M_2$  roots of multituberculates (**c**) (*Paulchoffatia*, GUI MAM 8/73). Images are not to scale. *Hahnodon* redrawn from previous publications<sup>14, 21</sup>. **a-b c**, anterobuccal cusp; **b f**, anterobuccal furrow; **I**, incisor position (homologous positions to earlier haramiyidans); **I f**, posterolingual furrow;  $M_2$ , second lower molar; **p-l c**, posterolingual cusp;  $PC^4$ , fourth postcanine (posterior molar).

Haramiyidans have frequently been included with multituberculates in ‘Allotheria’, a disputed taxonomic grouping that was diagnosed by dentition, with molars that have multiple cusp rows and palinal (anteroposterior) motion during mastication<sup>4, 6, 9</sup>. The limited overlapping morphology that is available in eleutherodonts and earlier haramiyidans (for example, *Thomasia* and *Haramiyavia*), coupled with a sparse fossil record, complicates our ability to resolve the group’s global radiation and purported relationships to multituberculates. Here, we report an exceptionally preserved cranium from the basal Cretaceous of Utah, USA, that fills substantial morphologic and spatiotemporal gaps in our understanding of this important group. The specimen provides the basis of a new genus and species.

Mammaliaformes sensu Rowe (1986)<sup>13</sup>  
Haramiyida Hahn, Sigogneau-Russell and Wouters (1989)<sup>2</sup>  
Hahnodontidae Sigogneau-Russell (1991)<sup>14</sup>  
*Cifelliodon* gen. nov.  
*Cifelliodon wahkarmoosuch* sp. nov. (Fig. 1)

**Etymology.** Cifelli’s tooth (Latin: -odon) of the Yellow Cat (Ute language: yellow, wahkar; cat, moosuch). Genus name honours Richard Cifelli for his contributions to Cretaceous mammal research in the American West.

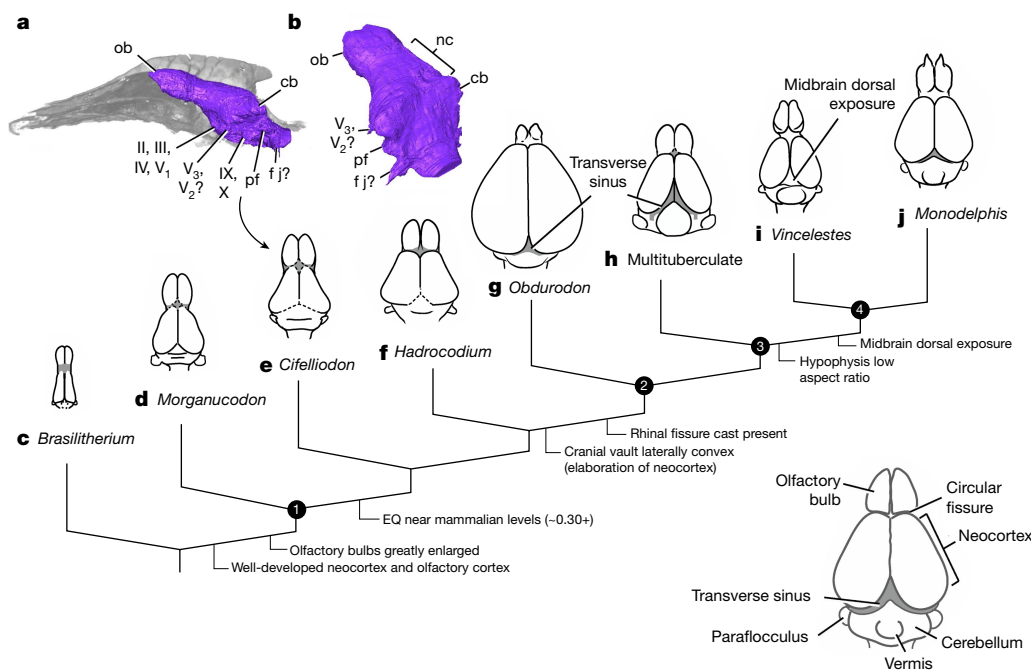
**Holotype.** An exceptionally preserved skull, UMNH VP 16771 (Natural History Museum of Utah, Vertebrate Paleontology Collection).

**Locality and horizon.** The holotype is from the Andrew’s Site’ quarry in the Lower Cretaceous Yellow Cat Member, Cedar Mountain Formation, Grand County, Utah, USA<sup>15</sup> (Extended Data Fig. 1). Radiometric dating places the age between approximately 139 and 124 million years old (see Methods and Supplementary Information).

**Diagnosis.** Medium-to-large Mesozoic mammaliaform with broad, shallow skull and rostrum and a reduced marginal tooth count; dental formula:  $I2:C1:PC4$ ; ultimate upper molars with high anterobuccal cusp and low, broad posterolingual cusp connected by a low ridge; septomaxilla absent; incisive foramina enlarged and positioned posteriorly on palate behind the level of the last (posterior) incisor pair; massive pterygoid transverse process that extends far ventral to the palatal surface; attenuated lacrimal anterior process with limited nasolacrimal contact; prominent sagittal crest; extensive occipital exposure of parietal and postparietal; plesiomorphic retention of a tabular bone; differs from *Hahnodon* in its larger size and higher aspect ratio of the rear molar in occlusal view (slightly more triangular than oval, with posterior apex).

**Description.** The skull of *Cifelliodon* (Fig. 1 and Extended Data Figs. 2–5) is large relative to other Mesozoic mammaliaforms (basal length around 70 mm; estimated body mass 0.91–1.27 kg)<sup>16, 17</sup> (see Methods and Supplementary Information). The upper dentition is largely missing, but preserves alveoli for two incisors, one canine and four postcanine maxillary teeth, including an in situ, unerupted ultimate molar pair (Fig. 2). The incisor roots were strongly posteriorly reclined and converged medially, suggesting procumbency as in haramiyidans, some multituberculates and gondwanatheres—another enigmatic group of mainly herbivorous mammal relatives that was restricted to Gondwana and has sometimes been classified as ‘allotheres’<sup>18, 19</sup>. The postcanine alveoli have approximately the same diameter as the canine alveolus and are also ‘undivided’ and reclined more posteriorly. The ultimate molar preserves a distinctive central valley that is closed-off anteriorly by a tall antero-buccal cusp and lingual ridge that connects to a lower posterolingual cusp, a pattern analogous to that of the fruit-eating hammer-headed bat *Hypsignathus monstrosus*<sup>20</sup>. Other haramiyidans shared a similar configuration with a chewing cycle that had both orthal and palinal movements<sup>4</sup>, but their upper molars differed in that the posterior cusp is larger than the anterior. In this regard, the ultimate molar of *Cifelliodon* is a closer analogue to *Hypsignathus* than that of other haramiyidans, and, except for its larger size and proportional differences, is nearly identical to an isolated tooth from the Lower Cretaceous of Morocco previously regarded as a multituberculate lower molar ( $M_2$ ): *Hahnodon taqueti*<sup>14, 21</sup>. Here, we regard the assignment of *Hahnodon* to multituberculates as problematic because, similar to *Cifelliodon*, the molar morphology includes a prominent antero-buccal cusp that is uncharacteristic of plagiaulacidan multituberculate  $M_2$ s, a central valley that is closed off anteriorly and incompletely divided roots (multituberculates display completely bifurcated roots; Fig. 2c). Other authors have also questioned the multituberculate affinity of hahnodontid teeth, tentatively suggesting that they be moved to Haramiyida<sup>5</sup>.

The skull shares additional features with some haramiyidans: upper incisor alveoli deep with procumbent orientation (Extended Data Figs. 6–8); lower canine fossa on maxilla-premaxilla absent; upper incisor number reduced to fewer than three; and upper postcanine loci reduced to five or fewer<sup>4, 9</sup> (Extended Data Table 1). Moreover, *Cifelliodon* retains



**Fig. 3 | Brain endocast in *Cifelliodon* and other mammaliaforms.**

**a, b**, Endocast of *Cifelliodon* in left lateral (**a**) and posterior oblique (**b**) views. **c**, Cynodont *Brasilitherium*<sup>29</sup>. **d**, Mammaliaform *Morganucodon*<sup>22</sup>. **e**, Hahnodontid mammaliaform *Cifelliodon*. **f**, *Hadrocodium*<sup>22</sup>. **g**, Monotreme *Obdurodon*<sup>22</sup>. **h**, Multituberculata (based on *Chulsanbaatar*)<sup>22</sup>. **i**, *Vincelestes*<sup>23</sup>. **j**, *Monodelphis*<sup>30</sup>. Drawings scaled to their relative encephalization quotients (EQ). Major nodes are: (1) Mammaliaformes,

(2) Mammalia, (3) Theriiformes, (4) Cladotheria (therians and close allies). cb, cerebellar hemisphere; f j, jugular foramen cast in perilymphatic fossa; nc, neocortex region; ob, olfactory bulb; pf, paraflocculus cast; II, III, IV, V<sub>1</sub>, cranial nerve exits near sphenorbital fissure; V<sub>3</sub>, V<sub>2</sub>, mandibular (and maxillary?) trigeminal branch roots; IX, X, glossopharyngeal and vagus nerve roots.

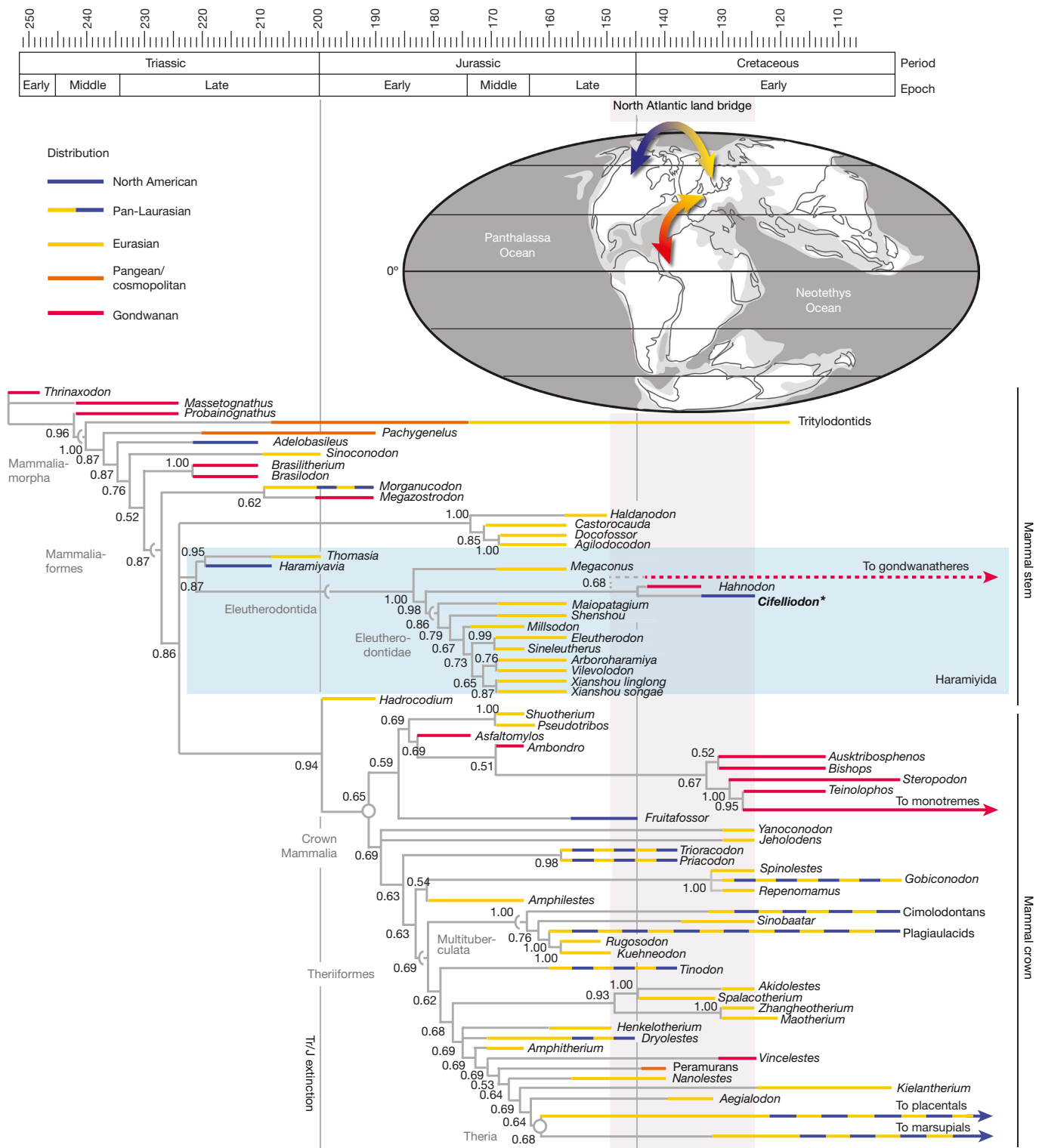
several plesiomorphic traits seen in Triassic mammaliaforms, but not exhibited widely by previously known Cretaceous mammals, including: parietal lateral margins waisted rather than bulging outward; tabular bone present (also in the gondwanatherian *Vintana*); pterygopalatal ridges present (also in select theriiforms, including cimolodontan multituberculates<sup>19</sup>); pterygoid flanges massive; secondary palate posterior limit well anterior to level of ultimate molar; maxilloturbinal supporting ridges absent on internal surface of maxilla; palatine anterodorsal expansion modest, such that maxilla and frontal maintain contact on anteromedial wall of orbit; and anterior part of the jugal extends to the facial part of the maxilla and forms a part of the anterior orbit. Nevertheless, the specimen also exhibits crown mammal features, including: a premaxilla facial process that is extensive and borders on the nasal; premaxilla with reduced internarial process; and posterior portion of jugal that terminates well anterior to the squamosal glenoid.

High-resolution X-ray computed tomography reveals an endocranial morphology that is transitional between earlier stem mammals and crown Mammalia (Fig. 3). The cranial vault is relatively modest in size (encephalization quotient 0.25–0.30; following a previously published method<sup>22</sup>), smaller than in the mammaliaform *Hadrocodium* (approximately 0.50) and some Cretaceous crown mammals (for example, *Vincelestes*, 0.37; *Pucadelphys*, 0.32)<sup>22, 23</sup>. Much of the enlargement in brain volume was formed by the massive olfactory bulbs and piriform cortex as in other early mammaliaforms. Given the medially waisted parietals and limited endocranial capacity, there was little expansion of the neocortex. The endocast surface is largely lissencephalic, olfactory bulbs massive and a shallow peduncle in place of the circular fissure, thus showing poor gross differentiation of neocortical structures. This contrasts with early therians and other crown mammals more typical of the Cretaceous that show greater differentiation of neocortical structures and their boundaries in endocasts (Fig. 3). Many posterior endocranial structures that are visible on the endocast (perilymphatic foramen cast, parafloccular cast) are located at about the level of the wide, shallow glenoid fossa as in other early mammaliaforms, but not as posteriorized as in large-brained therians. As in other early

mammaliaforms, the endocranium of *Cifelliodon* supports the premise that haramiyids exhibited a similar degree of encephalization and behavioural complexity that was driven mainly by olfaction<sup>22</sup>.

We assessed the phylogenetic position of *Cifelliodon* and Haramiyida using an updated morphological character matrix<sup>4, 24</sup>. Broad-level relationships support that Haramiyida is monophyletic and represents a close sister group to Mammalia<sup>4, 24</sup>, rather than a paraphyletic assemblage of crown mammals related to multituberculates (in contrast to Zheng et al.<sup>6</sup>) (Fig. 4 and Extended Data Fig. 9). We define Haramiyida as all mammaliaforms more closely related to *Thomasia* and *Haramiyavia* than to *Didelphis*. Within Haramiyida, *Cifelliodon* forms a relationship with *Hadrodon* in a polytomy with the gondwanatherian *Vintana*<sup>18</sup>. However, the position of gondwanatheres is only weakly supported; for example, an alternate tree topology constraining *Vintana* to a sister taxon relationship with multituberculates results in a tree score that is only nine steps longer than our most parsimonious tree (Templeton test,  $P = 0.139$ ). However, support for a monophyletic Haramiyida outside crown Mammalia is significantly more robust than moving the clade into crown mammals along with *Vintana* and multituberculates (as in the previously published topology<sup>18</sup>), which requires 24 additional steps ( $P = 0.032$ ).

Although the haramiyidan affinity of hahnodontids has been tenuous in the absence of more complete fossils, the exceptional preservation of *Cifelliodon* helps to strengthen the group's relationship to Haramiyida while revising its biogeographic history. *Cifelliodon* supports that haramiyidan stem mammals survived into the Early Cretaceous and sustained broad distributions in Laurasia and Gondwana, bridging Jurassic and Cretaceous terrestrial ecosystems across continents. The only previous haramiyidan material of Gondwanan origin included the isolated dentition of *Allostaffia* from the Jurassic of Tanzania<sup>25</sup> and *Avashishta*—an isolated molar of uncertain affinities from the Upper Cretaceous of India<sup>26</sup>. The further presence of hahnodontids in North America and Africa mirrors underappreciated Pangaea-wide occurrences of Cretaceous tetrapods, including several shared dinosaurian taxa<sup>27, 28</sup> that linked northern and southern continents much later than previously recognized (see Supplementary Information).



**Fig. 4 | Time-calibrated phylogeny and biogeography of Mesozoic mammalian and nonmammalian synapsid taxa.** Bayesian consensus cladogram for the analysis of 125 mammalian and nonmammalian synapsid taxa and 538 morphological characters. Note that haramiyidan stem mammals, previously restricted to Laurasia in the Jurassic period, exhibited distributions into Gondwana

by the Cretaceous period. Cretaceous nonmammalian mammalian morphs were rare, but globally widespread. Values at the nodes represent Bayesian posterior probabilities. Values on the time scale are shown as million years before present. North Atlantic land bridge interval is indicated with a grey bar (maximum duration). Tr/J, Triassic/Jurassic extinction.

Ultimately, the Northern Hemisphere radiation of multituberculates and tribosphenic mammals was offset by losses of many archaic mammalian morph groups during the mid-Cretaceous, including some of the longest-lived mammalian morph subclades: tritylodontids, docodonts, haramiyidans, plagiaulacidan multituberculates and reduced numbers of triconodonts and 'symmetrodonts'<sup>1</sup>. *Cifelliodon* supports the premise

that large-bodied mammalian morphs of diverse ecologies were present in North American communities up until this mid-Cretaceous diversity bottleneck. Our discovery of *Cifelliodon* and ongoing fieldwork in the Lower Cretaceous of Utah strengthen the hypothesis that corridors of Pangea-wide dispersal were accessible to relictual clades of large and small vertebrates at least as recently as the Early Cretaceous<sup>27, 28</sup>.

Consequently, the enigma of why some groups were better at dispersing than others (for example, the conspicuous absence of African multituberculates, an otherwise prolific clade) is more likely to be a consequence of collecting biases or unknown biotic factors rather than strictly tectonic controls or ocean barriers.

## Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0126-y>.

Received: 29 November 2017; Accepted: 10 April 2018;  
Published online 23 May 2018.

- Grossnickle, D. M. & Polly, P. D. Mammal disparity decreases during the Cretaceous angiosperm radiation. *Proc. R. Soc. B* **280**, 20132110 (2013).
- Hahn, G., Sigogneau-Russell, D. & Wouters, G. New data on Theriidae—their relations with Paulchoffatiidae and Haramiyidae. *Geol. Paleontol.* **23**, 205–215 (1989).
- Butler, P. M. Review of the early altherian mammals. *Acta Palaeontol. Pol.* **45**, 317–342 (2000).
- Luo, Z.-X., Gatesy, S. M., Jenkins, F. A. Jr, Amaral, W. W. & Shubin, N. H. Mandibular and dental characteristics of Late Triassic mammaliaform *Haramiyavia* and their ramifications for basal mammal evolution. *Proc. Natl Acad. Sci. USA* **112**, E7101–E7109 (2015).
- Butler, P. M. & Hooker, J. J. New teeth of altherian mammals from the English Bathonian, including the earliest multituberculates. *Acta Palaeontol. Pol.* **50**, 185–207 (2005).
- Zheng, X., Bi, S., Wang, X. & Meng, J. A new arboreal haramiyid shows the diversity of crown mammals in the Jurassic period. *Nature* **500**, 199–202 (2013).
- Zhou, C. F., Wu, S., Martin, T. & Luo, Z.-X. A Jurassic mammaliaform and the earliest mammalian evolutionary adaptations. *Nature* **500**, 163–167 (2013).
- Meng, J., Bi, S., Zheng, X. & Wang, X. Ear ossicle morphology of the Jurassic euharamiyid *Arboroharamiya* and evolution of mammalian middle ear. *J. Morphol.* **279**, 441–457 (2018).
- Bi, S., Wang, Y., Guan, J., Sheng, X. & Meng, J. Three new Jurassic euharamiyid species reinforce early divergence of mammals. *Nature* **514**, 579–584 (2014).
- Han, G., Mao, F., Bi, S., Wang, Y. & Meng, J. A Jurassic gliding euharamiyid mammal with an ear of five auditory bones. *Nature* **551**, 451–456 (2017).
- Luo, Z.-X. et al. New evidence for mammaliaform ear evolution and feeding adaptation in a Jurassic ecosystem. *Nature* **548**, 326–329 (2017).
- Meng, Q.-J. et al. New gliding mammaliaforms from the Jurassic. *Nature* **548**, 291–296 (2017).
- Rowe, T. B. *Osteological Diagnosis of Mammalia, L. 1758, and its Relationship to Extinct Synapsida*. PhD thesis, Univ. California, Berkeley (1986).
- Sigogneau-Russell, D. First evidence of Multituberculata (Mammalia) in the Mesozoic of Africa. *Neues Jahrb. Geol. Paläontol. Abh.* **1991**, 119–125 (1991).
- McDonald, A. T. et al. New basal iguanodonts from the Cedar Mountain formation of Utah and the evolution of thumb-spiked dinosaurs. *PLoS ONE* **5**, e14075 (2010).
- Gingerich, P. D. & Smith, B. H. in *Size and Scaling in Primate Biology* (ed. Jungers, W. L.) 257–272 (Plenum, New York, 1984).
- Millien, V. & Bovy, H. When teeth and bones disagree: body mass estimation of a giant extinct rodent. *J. Mamm.* **91**, 11–18 (2010).
- Krause, D. W. et al. First cranial remains of a gondwanatherian mammal reveal remarkable mosaicism. *Nature* **515**, 512–517 (2014).
- Krause, D. W., Wible, J. R., Hoffmann, S., Groenke, J. R., O'Connor, P. M., Holloway, W. L. & Rossie, J. B. Craniofacial morphology of *Vintana sertichi* (Mammalia, Gondwanatheria) from the Late Cretaceous of Madagascar. *J. Vertebr. Paleontol.* **34**, 14–109 (2014).
- Langevin, P. & Barclay, R. M. R. *Hypsignathus monstrosus*. *Mamm. Species* **357**, 1–4 (1990).
- Hahn, G. & Hahn, R. New multituberculate teeth from the Early Cretaceous of Morocco. *Acta Palaeontol. Pol.* **48**, 349–356 (2003).
- Rowe, T. B., Macrini, T. E. & Luo, Z.-X. Fossil evidence on origin of the mammalian brain. *Science* **332**, 955–957 (2011).
- Macrini, T. E., Rougier, G. W. & Rowe, T. Description of a cranial endocast from the fossil mammal *Vincelestes neuquenianus* (Therapsida) and its relevance to the evolution of endocranial characters in therians. *Anat. Rec. (Hoboken)* **290**, 875–892 (2007).
- Luo, Z.-X. et al. Evolutionary development in basal mammaliaforms as revealed by a docodontan. *Science* **347**, 760–764 (2015).
- Heinrich, W.-D. First haramiyid (Mammalia, Allotheria) from the Mesozoic of Gondwana. *Foss. Rec.* **2**, 159–170 (1999).
- Anantharaman, S., Wilson, G. P., Das Sarma, D. C. & Clemens, W. A. A possible Late Cretaceous “haramiyid” from India. *J. Vertebr. Paleontol.* **26**, 488–490 (2006).
- Brikatis, L. Late Mesozoic North Atlantic land bridges. *Earth Sci. Rev.* **159**, 47–57 (2016).
- Dunhill, A. M., Bestwick, J., Narey, H. & Sciberras, J. Dinosaur biogeographical structure and Mesozoic continental fragmentation: a network-based approach. *J. Biogeogr.* **43**, 1691–1704 (2016).
- Rodrigues, P. G., Ruf, I. & Schultz, C. L. Study of a digital cranial endocast of the non-mammaliaform cynodont *Brasilitherium riograndensis* (Late Triassic, Brazil) and its relevance to the evolution of the mammalian brain. *Palaontol. Z.* **88**, 329–352 (2014).
- Macrini, T. E., Rowe, T. & Vandeberg, J. L. Cranial endocasts from a growth series of *Monodelphis domestica* (Didelphidae, Marsupialia): a study of individual and ontogenetic variation. *J. Morphol.* **268**, 844–865 (2007).

**Acknowledgements** We thank R. Irmis, C. Levitt-Bussian and the Natural History Museum of Utah (UMNH); Utah Geological Survey, which funded the excavation and preparation of the specimen; A. R. C. Milner who discovered Andrew's Site (UMNH 1207); Utah Friends of Paleontology and volunteers who helped with excavation; Bureau of Land Management for permission (BLM permit #UTEX-05-031) and site supervision in 2005 under J. Cavin; J. Cavin for initial preparation and S. Madsen for final preparation of the specimen; T. Martin for the *Paulchoffatia* jaw from the Guimarota Mammals Collection (GUI MAM) illustrated in Fig. 2c; E. Hsu, S. Merchant and University of Utah Small Animal Imaging facility, Salt Lake City, for high-resolution X-ray computed tomography scanning; J. Lungmus and A. Neander for additional scanning at the University of Chicago PaleocT; E. Seiffert (USC) for helpful comments. The specimen is housed at UMNH, Salt Lake City.

**Reviewer information** *Nature* thanks S. Hoffmann and T. Martin for their contribution to the peer review of this work.

**Author contributions** J.L.K. conducted fieldwork and oversaw laboratory preparation of the specimen; A.K.H., Z.-X.L. and J.A.S. scanned the specimen using computed tomography; A.K.H. analysed the data and conducted the computed tomography reconstructions; A.K.H. illustrated the specimen, and A.K.H. and Z.-X.L. composed the figures; A.K.H., D.M.G. and Z.-X.L. conducted the phylogenetic analysis; and A.K.H., D.M.G., J.L.K., J.A.S. and Z.-X.L. wrote the manuscript.

**Competing interests** The authors declare no competing interests.

## Additional information

**Extended data** is available for this paper at <https://doi.org/10.1038/s41586-018-0126-y>.

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41586-018-0126-y>.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>.

**Correspondence and requests for materials** should be addressed to A.K.H. or Z.-X.L.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## METHODS

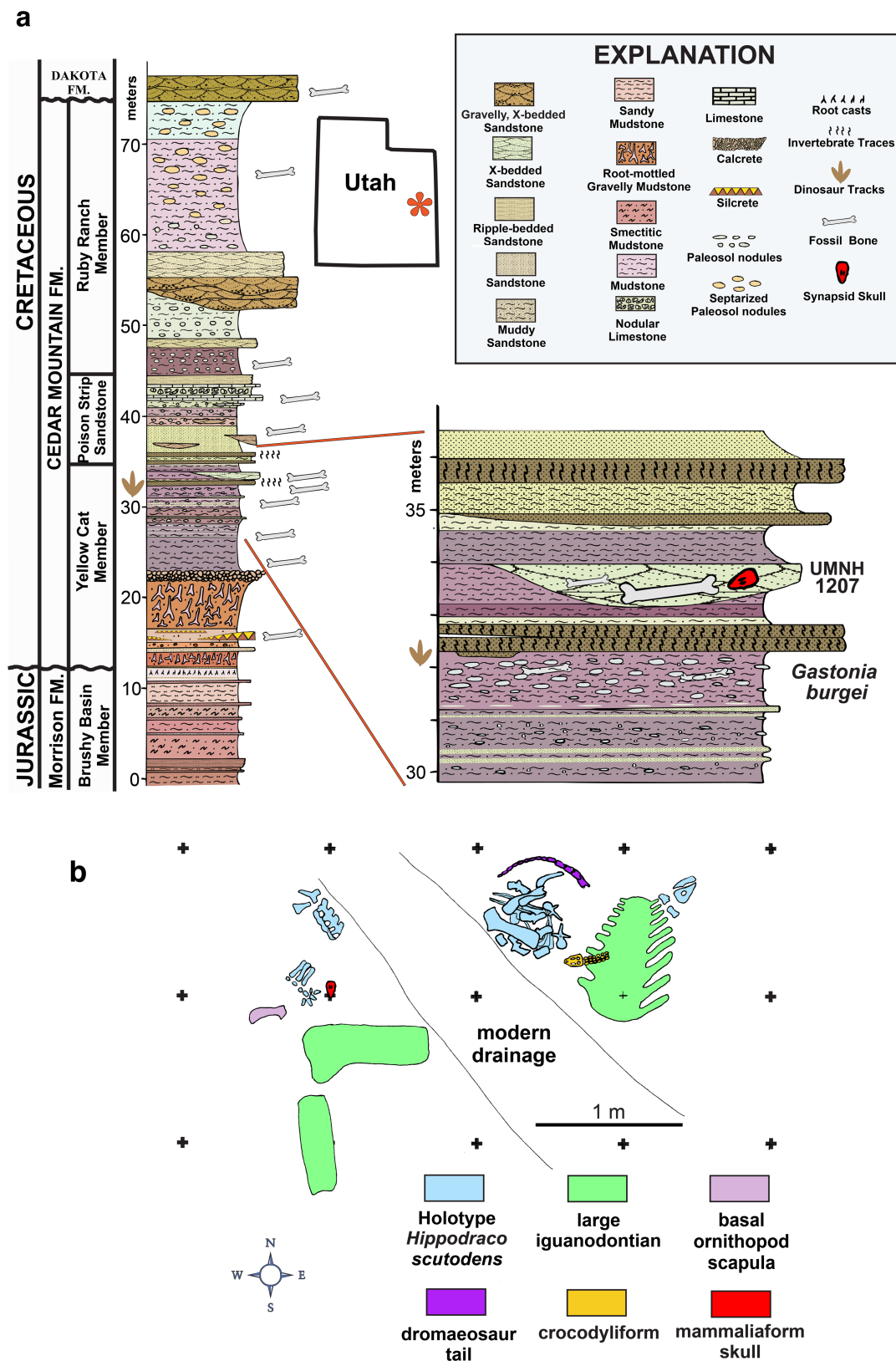
**High-resolution X-ray computed tomography.** The specimen was micro-CT scanned at the University of Utah's Core Small Animal Imaging facility using a Siemens INVEON micro-CT scanner (aluminium filter and energy settings of 180 kV and 500  $\mu$ A with a voxel size of 94.2  $\mu$ m<sup>3</sup>). Additional focused scanning was performed at University of Chicago's PaleoCT facility on a custom General Electric Phoenix VTOMEX S 240 to visualize the unerupted ultimate molars (energy settings of 180 kV and 500  $\mu$ A with a voxel size of 25.3  $\mu$ m<sup>3</sup>). Resulting DICOMs were extracted as image stacks and rendered as 3D models in Avizo 7.0 (Visage Imaging, San Diego, CA).

**Phylogenetic methods.** The previously published matrix<sup>4,24</sup> was modified to evaluate the relationships between *Cifelliodon* and other Mesozoic mammalian morphs. In total, 538 morphological characters from 125 mammal and nonmammalian synapsid taxa were analysed using Bayesian inference and maximum parsimony. Characters 1–515 correspond to those from the previously published studies<sup>4,24</sup>, characters 516–518 are borrowed from previously published data<sup>18</sup> and characters 519–538 are new. All characters had equal weight and none were ordered. We performed the Bayesian analysis in MrBayes version 3.2.6<sup>31</sup> implementing the standard Mk model for morphological evolution<sup>32</sup> with variable character rates. A log-normal rate distribution was chosen for the evolutionary model because of evidence, which suggested that log-normal distributions may be preferred for morphological datasets<sup>33,34</sup>. We ran the analysis for ten million generations (with the first 25% removed as burn-in) and sampled the posterior distribution every 1,000 generations. To increase the chance of state swapping between chains and to help to reduce the standard deviation of split frequencies (that is, help the runs to converge), the difference in 'temperature' between chains was reduced to 0.06 (from the default setting of 0.10). Parsimony analysis was performed in TNT version 1.1 (tree analysis using new technology)<sup>35</sup>. A heuristic search (random addition sequence with 1,000 replicates) was performed and recovered 1,920 equally most parsimonious trees (tree score = 2,776 steps; consistency index = 0.317; retention index = 0.795). The resulting Bayesian consensus and strict consensus parsimony trees are shown in Extended Data Fig. 9.

**Reporting summary.** Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

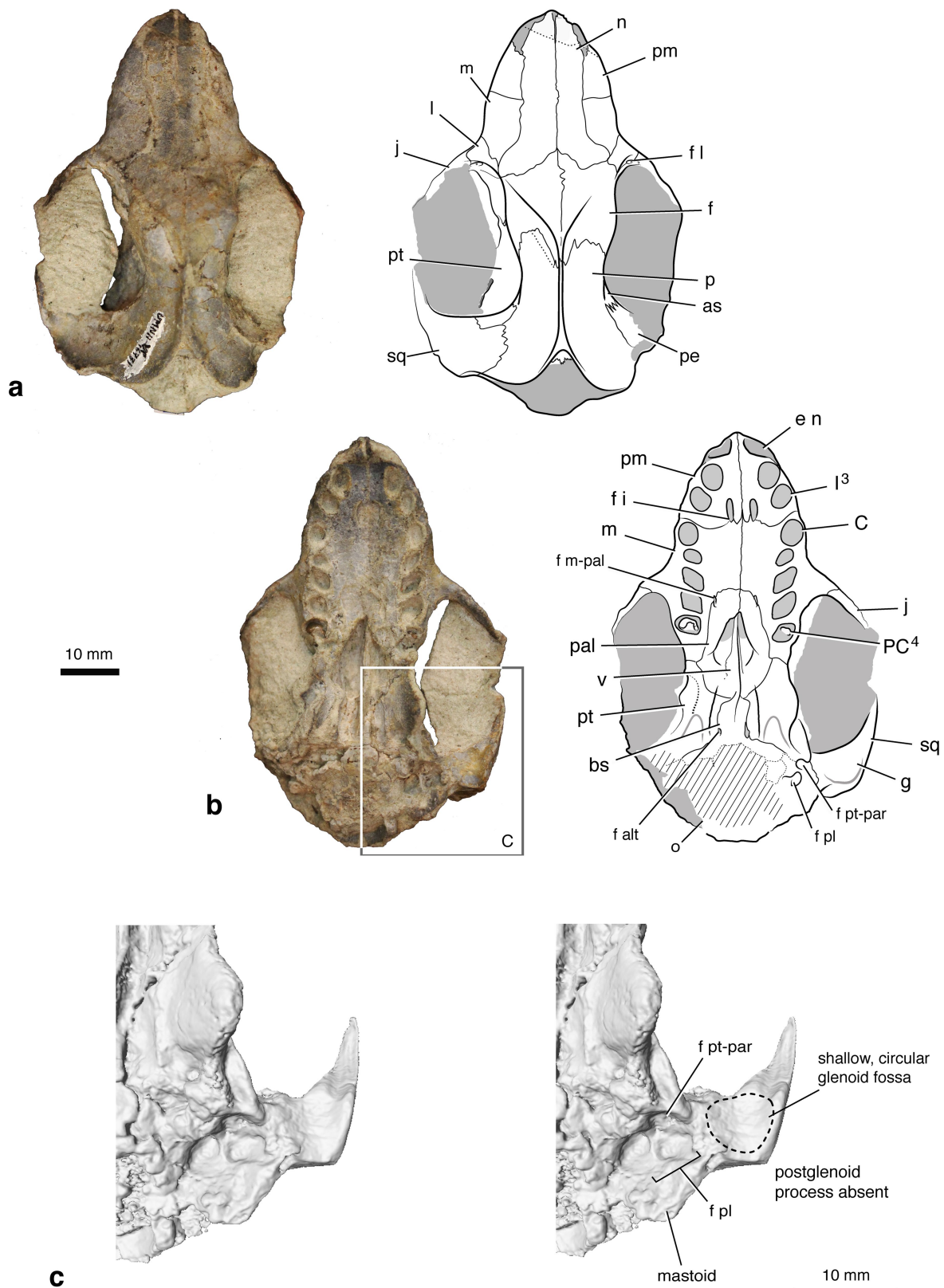
**Data availability.** The specimen is deposited at the UMNH, Salt Lake City, USA, and additional data are deposited online at <https://www.morphosource.org/>. CT image stacks are available from the corresponding authors upon request. Life Science Identifier (LSID): the new genus and species are registered with Zoobank (<http://zoobank.org>): urn:lsid:zoobank.org:act:C9D9F344-E058-4E7F-AD58-BAED111F574E.

31. Ronquist, F. & Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574 (2003).
32. Lewis, P. O. A likelihood approach to estimating phylogeny from discrete morphological character data. *Syst. Biol.* **50**, 913–925 (2001).
33. Wagner, P. J. Modelling rate distributions using character compatibility: implications for morphological evolution among fossil invertebrates. *Biol. Lett.* **8**, 143–146 (2012).
34. Harrison, L. B. & Larsson, H. C. Among-character rate variation distributions in phylogenetic analysis of discrete morphological characters. *Syst. Biol.* **64**, 307–324 (2015).
35. Goloboff, P. A., Farris, J. S. & Nixon, K. C. TNT, a free program for phylogenetic analysis. *Cladistics* **24**, 774–786 (2008).
36. Hoffmann, S., O'Connor, P. M., Kirk, E. C., Wible, J. R. & Krause, D. W. Endocranial and inner ear morphology of *Vintana sertichi* (Mammalia, Gondwanatheria) from the Late Cretaceous of Madagascar. *J. Vertebr. Paleontol.* **34**, 110–137 (2014).
37. Ruf, I., Luo, Z.-X. & Martin, T. Reinvestigation of the basicranium of *Haldanodon exspectatus* (Mammaliaformes, Docodonta). *J. Vertebr. Paleontol.* **33**, 382–400 (2013).
38. Hahn, G. & Hahn, R. in *Guimarota: A Jurassic Ecosystem* (eds. Martin, T. & Krebs, B.) 97–108 (Dr. Friedrich Pfeil, München, 2000).
39. Kermack, K. A., Mussett, F. & Rigney, H. W. The skull of *Morganucodon*. *Zool. J. Linn. Soc.* **71**, 1–158 (1981).
40. Kielan-Jaworowska, Z., Cifelli, R. L. & Luo, Z.-X. *Mammals from the Age of Dinosaurs: Origins, Evolution, and Structure* (Columbia Univ. Press, New York, 2004).



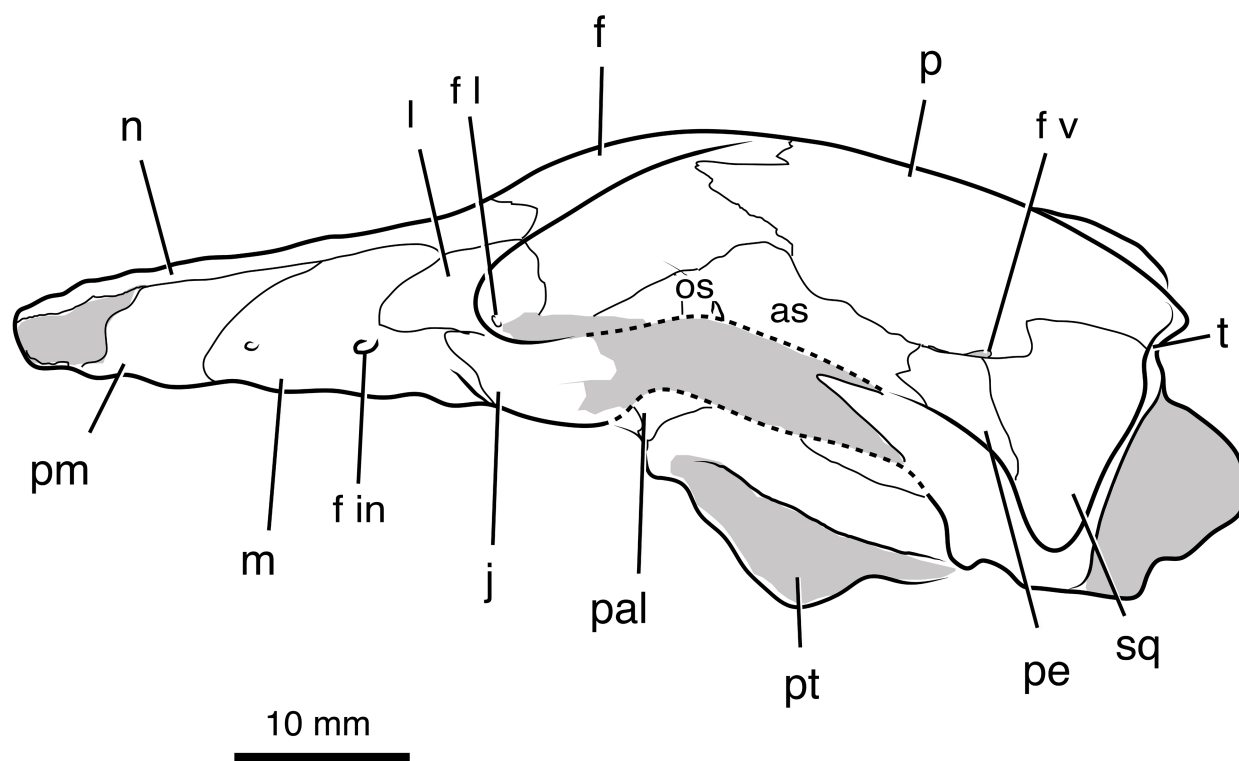
**Extended Data Fig. 1 | Stratigraphic section through Andrew's Site and quarry map. a,** Revised type section of the Yellow Cat Member of the Cedar Mountain Formation showing the stratigraphic position of

Andrew's Site (UMNH 1207). Modified from McDonald et al.<sup>15</sup>. **b,** Quarry map of Andrew's Site. The position of the mammaliaform skull UMNH VP 16771 is indicated in red.



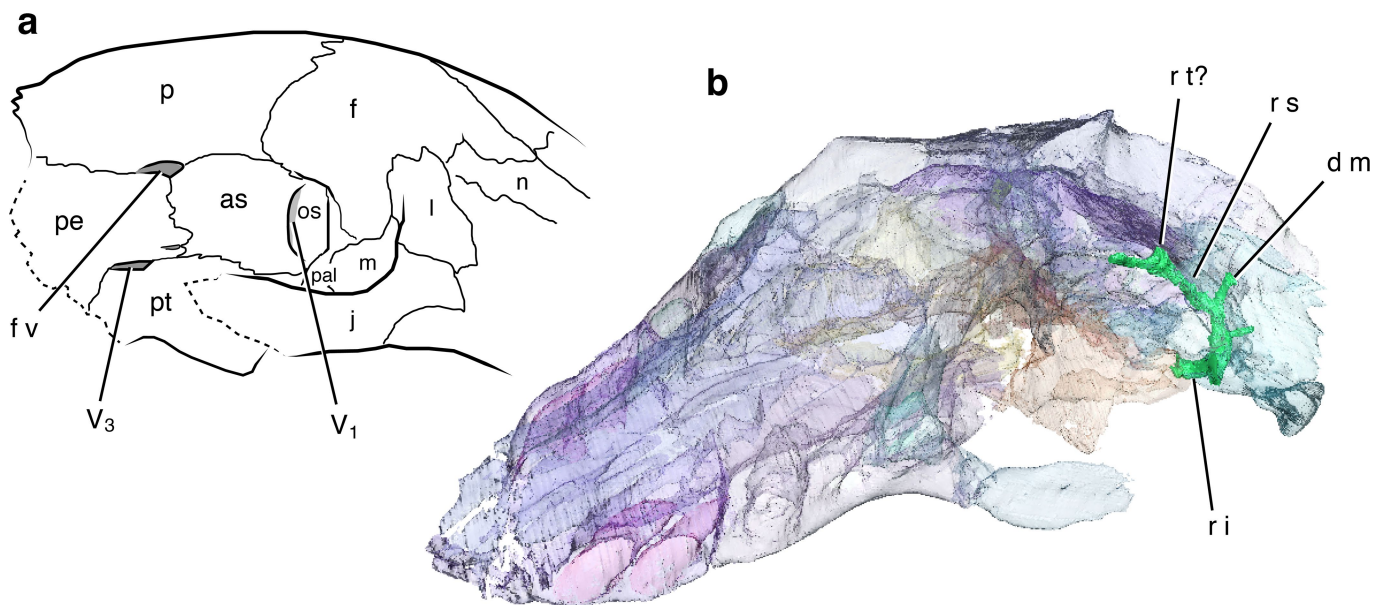
**Extended Data Fig. 2 | Skull and interpretive drawings of *C. wahkarmoosuch* UMNH VP 16771. a, b, Specimen shown in dorsal (a) and ventral (b) views. Grey areas represent matrix infill, hatched areas represent crushed/desiccated bones, dashed lines represent cracks through the specimen. c, Stereopair of the left side of skull in ventral view to show broad, shallow squamosal glenoid. as, alisphenoid (epipterygoid); bs, basisphenoid; C, upper canine alveolus; e n, external naris; f, frontal; f alt,**

anterior lateral trough foramen (cavum epipticum ventral opening); f i, incisive foramen; f l, lacrimal foramen; f m-pal, maxillopalatine foramen; f pl, perilymphatic groove/foramen; f pt-par, pterygoparoccipital foramen; g, squamosal glenoid; l<sup>3</sup>, second upper incisor alveolus; j, jugal; l, lacrimal; m, maxilla; n, nasal; o, occipital; p, parietal; pal, palatine; PC<sup>4</sup>, in situ posterior upper postcanine (molar); pe, petrosal; pm, premaxilla; pt, pterygoid; sq, squamosal; v, vomer.



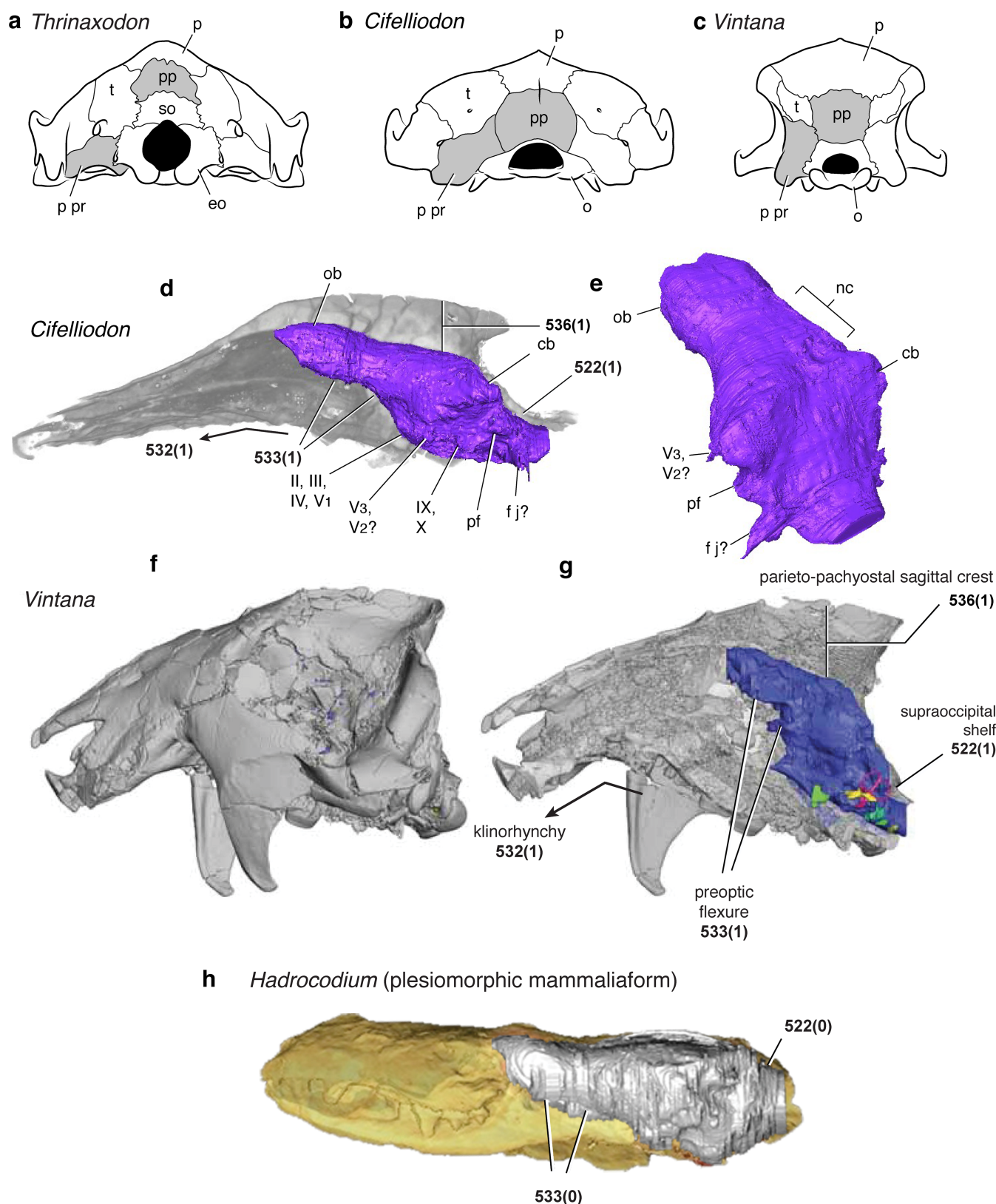
**Extended Data Fig. 3 | Skull and interpretive drawings of *C. wahkarmoosuch* UMNH VP 16771 (continued).** Specimen shown in left lateral view. Grey areas represent matrix infill, dashed lines represent

broken/reconstructed bones. f in, infraorbital foramen; f v, lateral external vascular foramen; os, orbitosphenoid; t, tabular.



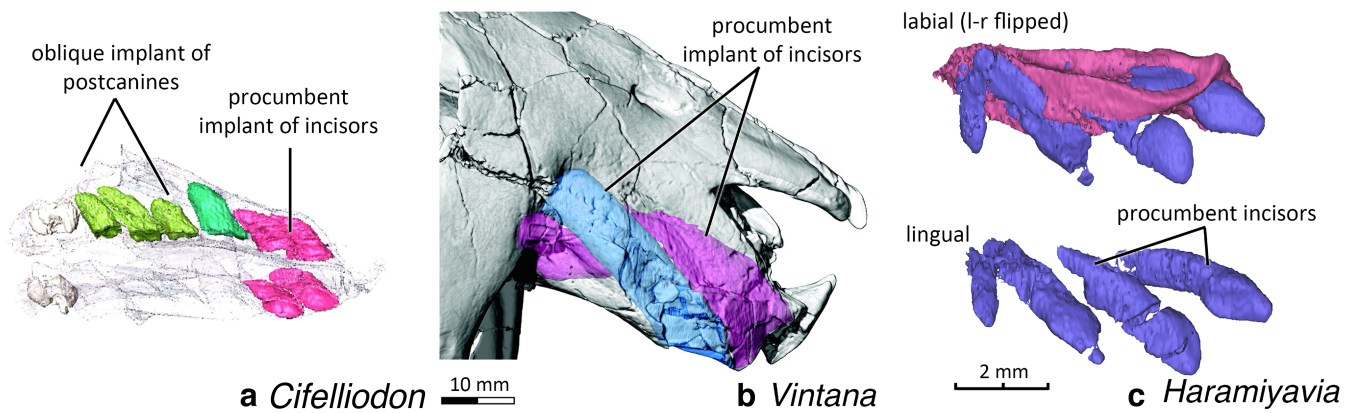
**Extended Data Fig. 4 | Cranial foramina and sinuses in the braincase of *C. wahkarmoosuch*.** **a**, Line drawing of the posterior skull in right oblique (anterolateral) view. **b**, Computed tomography transparency of the skull in left oblique view, showing the prootic sinus and associated branches of the stapedial artery (light green). d m, diploetica magna; r i, stapedial

artery (ramus inferior); r s, stapedial artery (ramus superior); r t, ramus temporalis; V<sub>1</sub>, orbital vacuity (hypothesized exit for ophthalmic branch of trigeminal nerve); V<sub>3</sub>, foramen pseudovale (hypothesized exit for mandibular branch of trigeminal nerve).



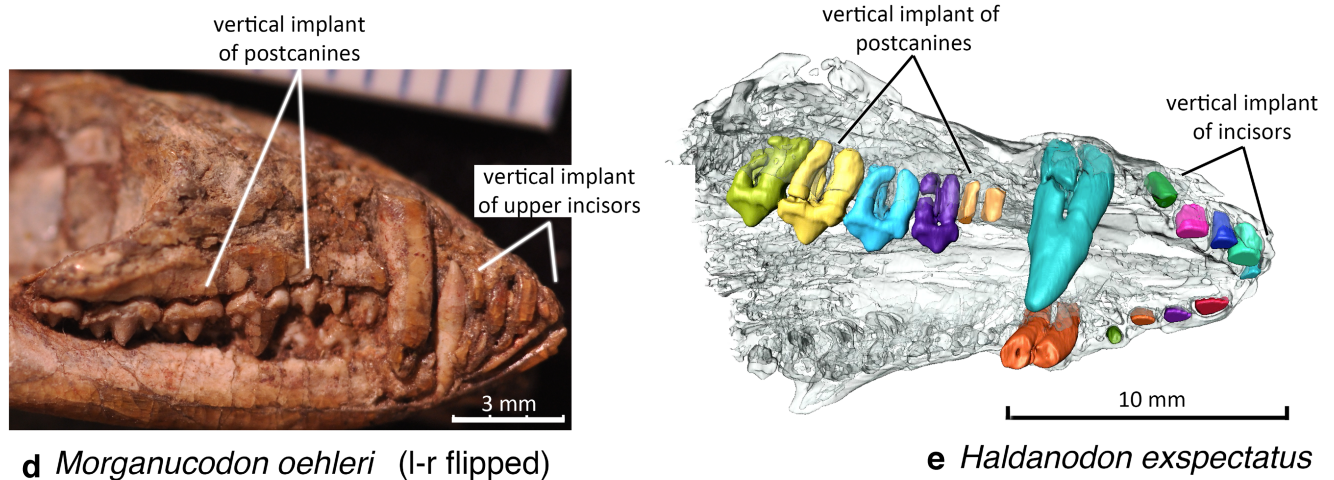
**Extended Data Fig. 5 | Cranial vault and endocranial features of *Cifelliodon*.** **a–c**, Comparison of occiputs of the cynodont *Thrinaxodon liorhinus* (based in part on University of California Museum of Paleontology (UCMP) 40466) (**a**), *C. wahkarmoosuch* (**b**) and *Vintana sertichi* (based on figure 1 of Krause et al.<sup>18</sup>) (**c**). Shaded bones in **b** and **c** emphasize sutural union of postparietal and paroccipital process. Skulls are not shown to scale. **d–h**, Brain endocasts. *Cifelliodon* endocast shown in left lateral (**d**) and posterior oblique (**e**) views. Skull of *Vintana* in lateral view shown externally (**f**) and with endocranial features exposed (**g**) (modified from figure 1 of Hoffmann et al.<sup>36</sup> flipped for comparison).

Skull of *Hadrocodium* (**h**) shown as a representative plesiomorphic mammaliaform (modified from figure 1 of Rowe et al.<sup>22</sup>; flipped for comparison). Numbers represent characters and character states described in the Supplementary Information. cb, cerebellar hemisphere; eo, exoccipital; f j, cast of jugular foramen within perilymphatic fossa; nc, region of neocortex; ob, olfactory bulb; p pr, paroccipital process; pf, paraflocculus cast; pp, postparietal; so, supraoccipital; II, III, IV, V<sub>1</sub>, exit for anterior cranial nerves near sphenorbital fissure; V<sub>3</sub>, V<sub>2</sub>, roots of mandibular (and maxillary?) branch of trigeminal nerve; IX, X, roots of glossopharyngeal and vagus nerves.



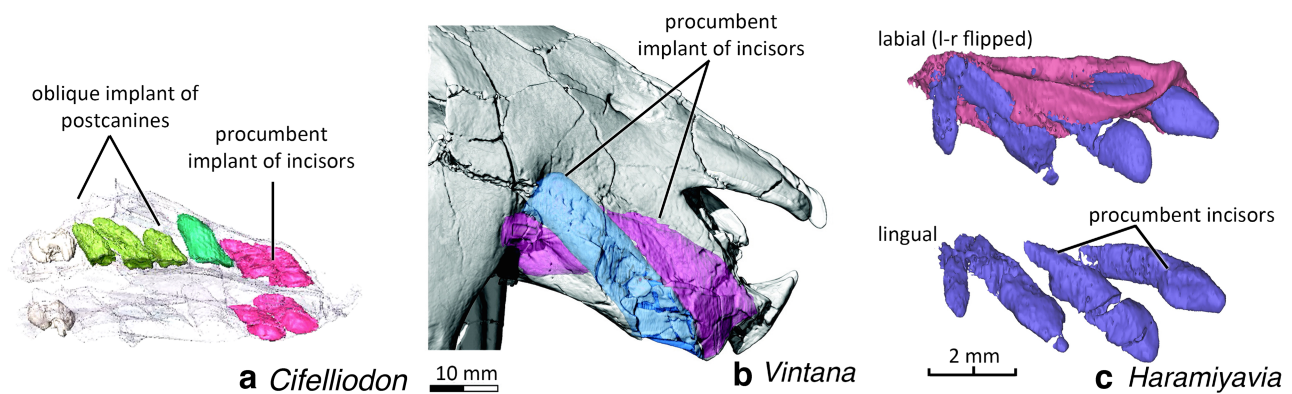
### Derived Haramiyidan Condition

### Typical Mammaliaform Condition



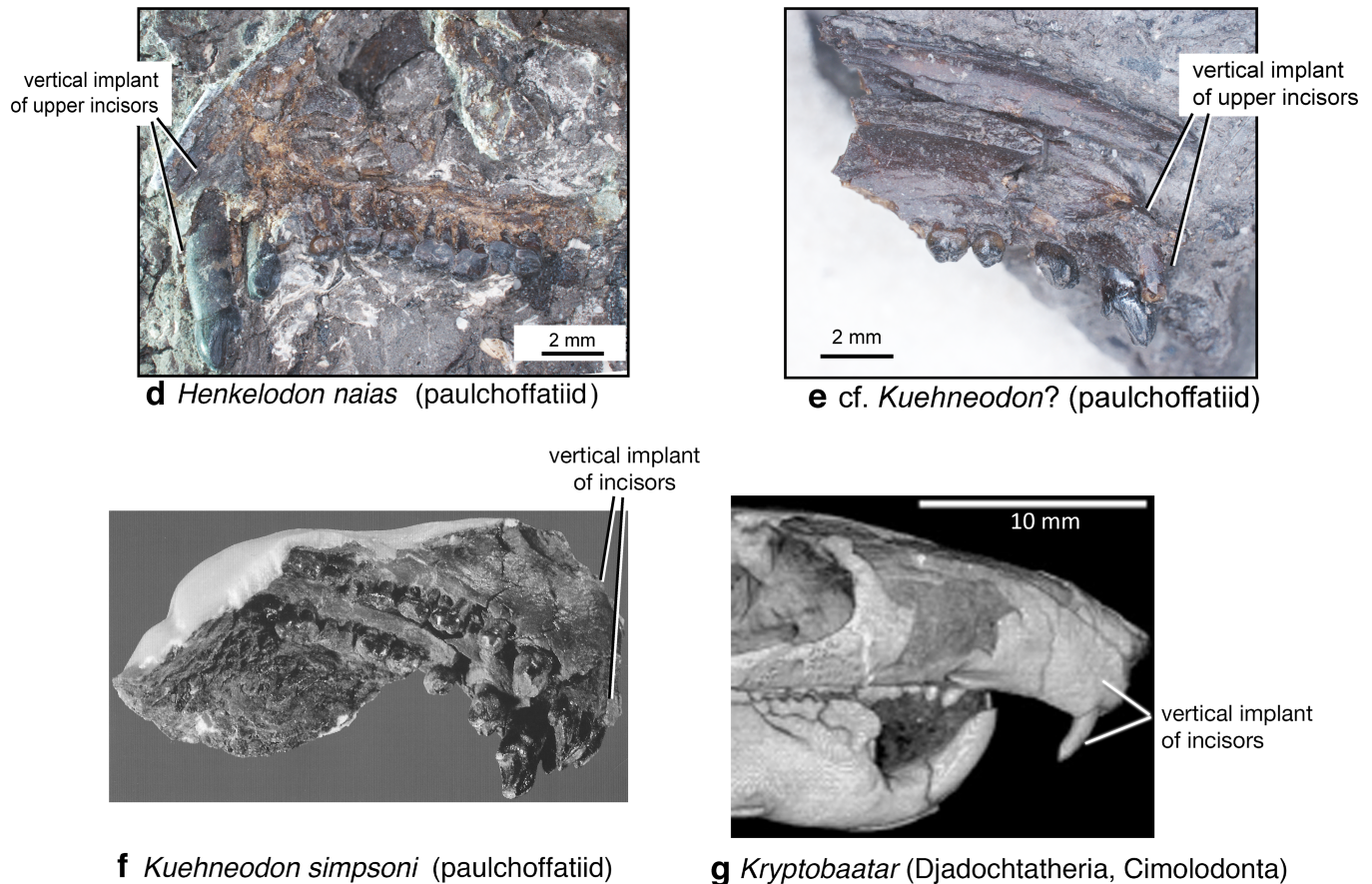
**Extended Data Fig. 6 | Derived condition of *Cifelliodon* and other haramiyidans in root morphology and root implant of upper teeth, in contrast to the primitive condition of mammaliaforms.** **a**, *C. wahkarmoosuch* (UMNH VP16771): procumbent implant of incisor roots. **b**, The Madagascar mammaliaform *Vintana* with procumbent incisor roots (based on figure 2 of Krause et al.<sup>18</sup>). **c**, *Haramiyavia*: the first three incisors are procumbent with elongate and procumbent roots (top-labial view, left–right flipped left premaxillary with in situ incisors; bottom-lingual view of left incisors; based on supplementary figure 1 of

Luo et al.<sup>4</sup>). **d**, *Morganucodon oehleri* (type specimen, photograph was made by Z.-X.L.): all upper teeth (including incisors) are vertical. **e**, *Haldanodon expectatus*: upper teeth are all vertical, orientation of incisors shown by their vertical alveoli (resegmented from the previously published dataset<sup>37</sup>). The typical condition for implant of upper teeth is vertical for most mammaliaforms. By comparison the primitive condition of most mammaliaforms, the procumbent incisor implant is a derived condition for haramiyidans, either on all of the incisors (*Cifelliodon* and *Vintana*) or on anterior incisors (*Haramiyavia*).



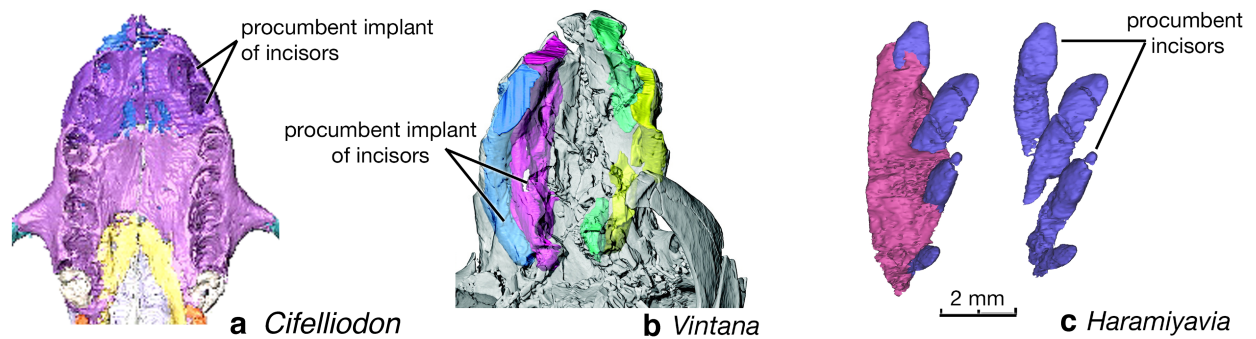
### Derived Haramiyidan Condition

### Typical Multituberculate Condition



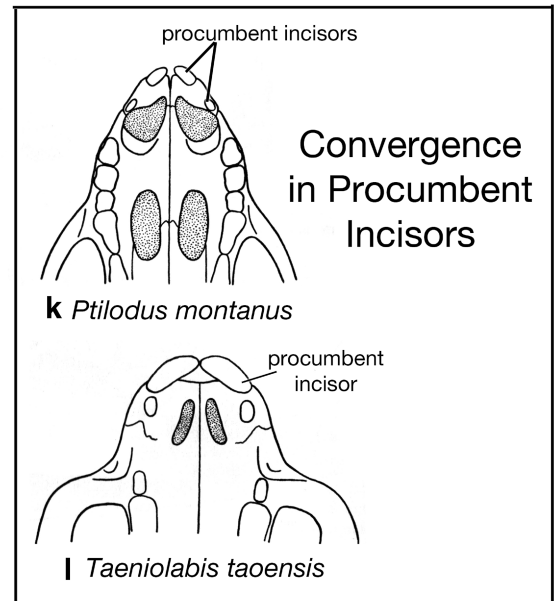
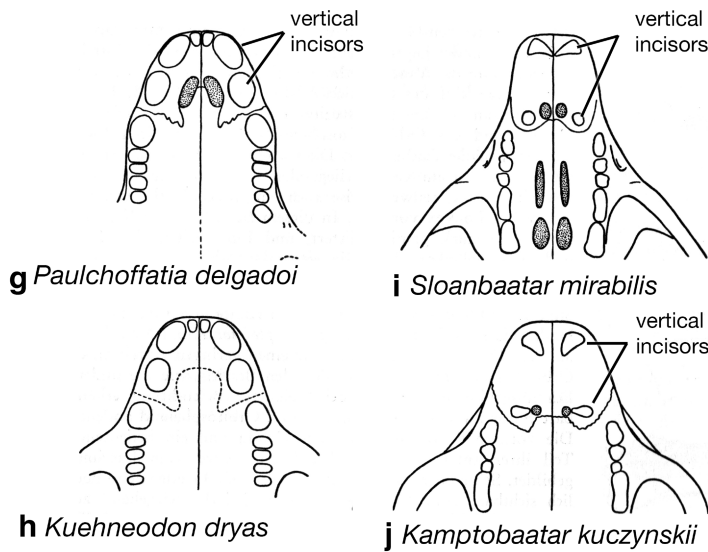
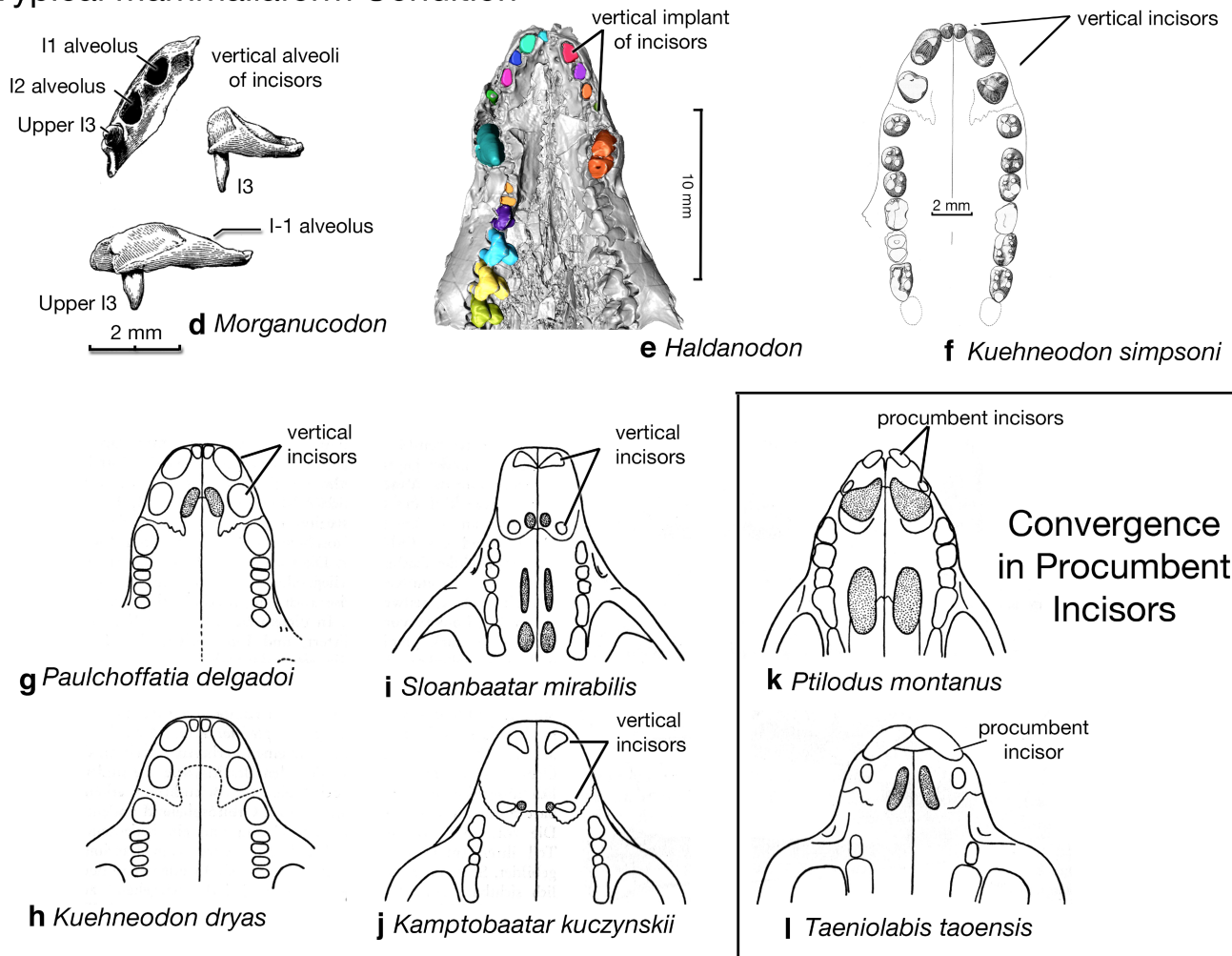
**Extended Data Fig. 7 | Derived condition of *Cifelliodon* and other haramiyidans in orientation and implant of incisor and postcanine roots, in contrast to the typical condition of multituberculates (in labial or lingual views).** **a.** *C. wahkarmoosuch* (UMNH VP16771): procumbent implant of incisor roots. **b.** *V. sertichi* with procumbent (hypertrophied) incisor roots (based on figure 2 of Krause et al.<sup>18</sup>). **c.** *Haramiyavia*: the first three incisors are procumbent with elongate and procumbent roots (top-labial view, left–right flipped left premaxillary with in situ incisors; bottom-lingual view of left incisors; based on supplementary figure 1 of Luo et al.<sup>4</sup>). **d.** *Henkelodon naias* (Paulchoffatiidae, Multituberculata: GUI-MAM-28-74). **e.** *Kuehneodon dryas* (Paulchoffatiidae, Multituberculata:

VJ400-155). **f.** *Kuehneodon simpsoni* (Paulchoffatiidae, Multituberculata; based on figure 12.5 from Hahn & Hahn<sup>38</sup>, reproduced with permission). **g.** *Kryptobaatar dashzevegi* (Djadochtatheria; Multituberculata; image from <http://digimorph.org/>). The procumbent orientation and tooth implant of incisors is a derived condition of haramiyidans (including *Cifelliodon*), in contrast to the vertically implanted incisors and postcanines of most Mesozoic multituberculates, including basal paulchoffatiids, plagiaulacids and djadochtatherians. The vertical implant of incisors in multituberculate condition is plesiomorphic, as it is also shared by other stem mammaliaforms.



## Derived Haramiyidan Condition

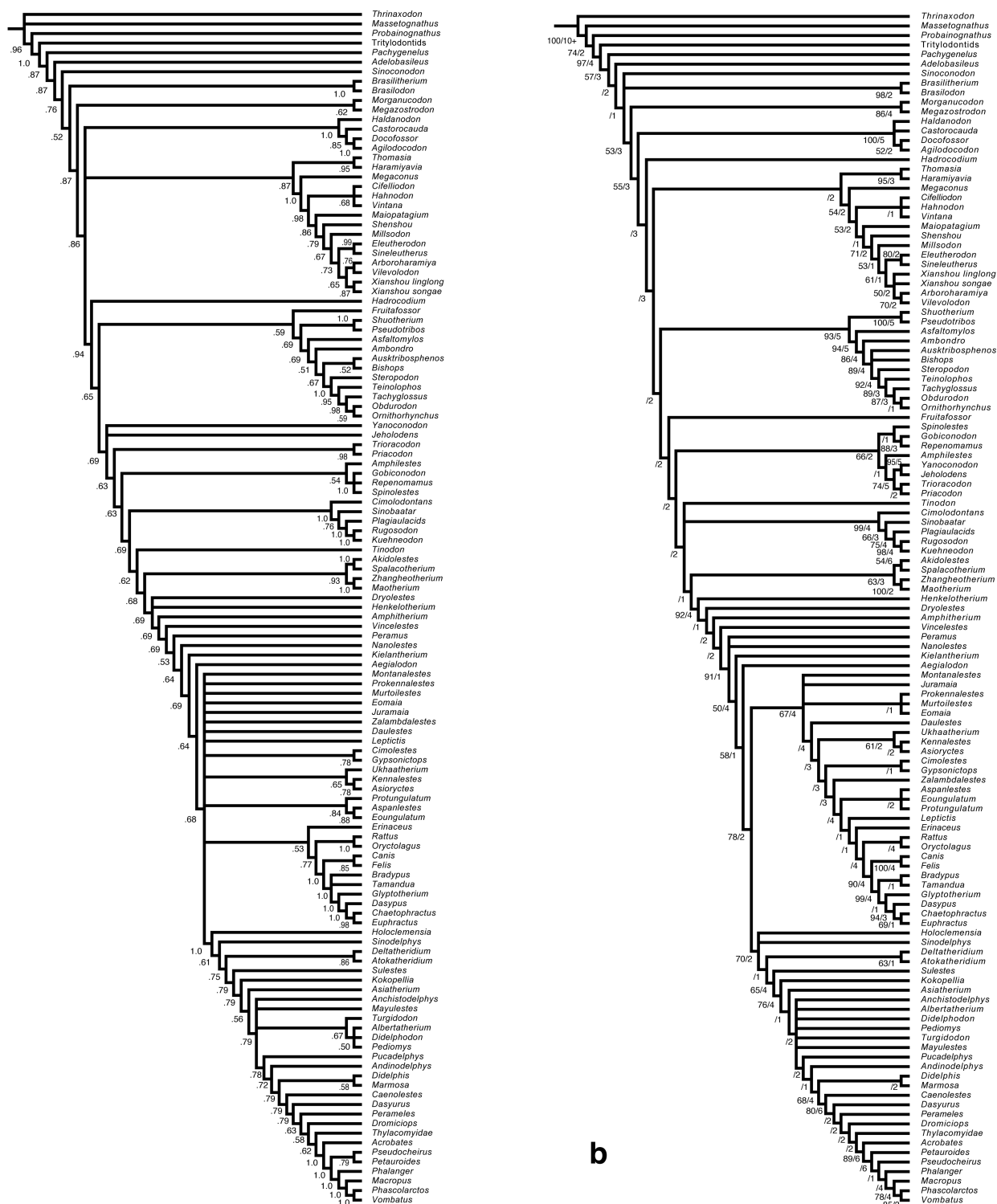
### Typical Mammaliaform Condition



## Convergence in Procumbent Incisors

**Extended Data Fig. 8 | Derived condition of *Cifelliodon* and other haramiyidan in orientation and implant of incisor roots, in contrast to the typical condition of other stem mammaliaforms and most multituberculates (palatal view). a, *C. wahkarmoosuch* (UMNH VP16771): procumbent implant of incisor roots. b, *V. sertichi* with procumbent (hypertrophied) incisor roots (based on figure 2 of Krause et al.<sup>18</sup>). c, *Haramiyavia*: the first three incisors are procumbent with elongate and procumbent roots (ventral view of left incisors and premaxillary: left with premaxillary; right premaxillary rendered invisible: based on supplementary figure 1 of Luo et al.<sup>4</sup>). d, *Morganucodon watsoni* (based on figure 6 of Kermack et al.<sup>39</sup>). e, *Haldanodon exspectatus*: vertical orientation of incisors shown by their vertical alveoli (resegmented from the previously published dataset<sup>37</sup>). f, *K. simpsoni* (Paulchoffatiidae, Multituberculata) (adapted with permission from Kielan-Jaworowska et al.<sup>40</sup>). g, *Paulchoffatia delgadoi* (Paulchoffatiidae, Multituberculata) (adapted with permission from Kielan-Jaworowska et al.<sup>40</sup>). h, *K. dryas* (Paulchoffatiidae, Multituberculata) (adapted with permission from Kielan-Jaworowska et al.<sup>40</sup>). i, *Sloanbaatar mirabilis* (Djadochtatheria,**

Multituberculata) (adapted with permission from Kielan-Jaworowska et al.<sup>40</sup>). j, *Kamptobaatar kuczynskii* (Djadochtatheria, Multituberculata) (adapted with permission from Kielan-Jaworowska et al.<sup>40</sup>). k, *Ptilodus montanus* (adapted with permission from Kielan-Jaworowska et al.<sup>40</sup>). l, *Taeniolabis taoensis* (adapted with permission from Kielan-Jaworowska et al.<sup>40</sup>). The procumbent implant of upper incisors is a derived condition for haramiyidans, either on all of the incisors (*Cifelliodon* and *Vintana*) or on anterior incisors (*Haramiyavia*). By comparison, the plesiomorphic condition of incisor implantation is vertical for most stem mammaliaforms (for example, *Morganucodon* and *Haldanodon*). All Mesozoic multituberculates have vertical implant of upper incisors, including basal-most paulchoffatiids that are preserved with upper incisors, plagiaulacids and djadochtatherians. However, the Cenozoic ptilodontid and taeniolabidid multituberculates show procumbent incisors. Because Mesozoic multituberculates have vertical implantation of incisors, the procumbent incisors of ptilodontids and taeniolabidids are interpreted to be secondarily derived and convergent (k, l).



**Extended Data Fig. 9 | Phylogenetic results. a**, Consensus cladogram from Bayesian analysis in MrBayes 3.2<sup>31</sup>. Branches are not drawn proportional to lengths. Clade support values shown at the nodes are Bayesian posterior probabilities (0.50–1.0). **b**, Strict consensus of 1,920 equally parsimonious trees (tree score = 2,776 steps; consistency

index = 0.317; retention index = 0.795) from the parsimony analysis in TNT 1.1<sup>35</sup>. Values at nodes are bootstrap values above 50% (first number (before the solidus)) and Bremer indices (second number (after the solidus)).

**Extended Data Table 1** | Major haramiyidan craniodental characters recognized in the present study

Character	Haramiyidae*	<i>Cifelliodon</i> (hahnodontid)	<i>Vintana</i> (gondwanathere)	Eleutherodontidae‡
<b>Dentition</b>				
occlusal action	orthal and palinal	orthal and palinal	mostly palinal	orthal and palinal
incisor orientation	marked procumbence	marked procumbence	marked procumbence	marked procumbence
upper incisor number	up to four	only two	only two	two or fewer
upper postcanine loci	?	four	up to five	five or fewer
upper premolar number	two or fewer	two or fewer	two or fewer	two or fewer
upper tooth row mesiolateral divergence	?	present	present	?
<b>Skull table</b>				
klinorhynch	?	present	present	present
tabular bone	?	present†	present†	?
lower canine fossa	?	absent	absent	absent
petrosal-postparietal sutural contact	?	present	present	?
supraoccipital shelf	?	present	present	?
pachyostotic parietal crest	?	present	present	?
<b>Endocranium</b>				
ossified cribiform plate	?	absent†	present?	?
ossified maxilloturbinal bones	?	absent†	absent?†	?
preoptic flexure	?	present	present	?

\* Including *Thomasia* and.

† Plesiomorphic retention.

‡ Including *Eleutherodon* and *Sineleutherus*, as well as the closely allied *Arboroharamiya*, *Shenshou* and *Xianshou*; 'Euharamiyida' of Bi et al.<sup>9</sup>.

# Body-size shifts in aquatic and terrestrial urban communities

Thomas Merckx<sup>1\*</sup>, Caroline Souffreau<sup>2</sup>, Aurélien Kaiser<sup>1</sup>, Lisa F. Baardsen<sup>3</sup>, Thierry Backeljau<sup>3,4</sup>, Dries Bonte<sup>5</sup>, Kristien I. Brans<sup>2</sup>, Marie Cours<sup>6</sup>, Maxime Dahirel<sup>5,7</sup>, Nicolas Debortoli<sup>8</sup>, Katrien De Wolf<sup>4</sup>, Jessie M. T. Engelen<sup>2</sup>, Diego Fontaneto<sup>9</sup>, Andros T. Gianuca<sup>2,10,11</sup>, Lynn Govaert<sup>2</sup>, Frederik Hendrickx<sup>4,5</sup>, Janet Higuti<sup>12</sup>, Luc Lens<sup>5</sup>, Koen Martens<sup>6,13</sup>, Hans Matheve<sup>5</sup>, Erik Matthysen<sup>3</sup>, Elena Piano<sup>4,14</sup>, Rose Sablon<sup>4</sup>, Isa Schön<sup>6,15</sup>, Karine Van Doninck<sup>8</sup>, Luc De Meester<sup>2,16</sup> & Hans Van Dyck<sup>1,16</sup>

**Body size is intrinsically linked to metabolic rate and life-history traits, and is a crucial determinant of food webs and community dynamics<sup>1,2</sup>. The increased temperatures associated with the urban-heat-island effect result in increased metabolic costs and are expected to drive shifts to smaller body sizes<sup>3</sup>. Urban environments are, however, also characterized by substantial habitat fragmentation<sup>4</sup>, which favours mobile species. Here, using a replicated, spatially nested sampling design across ten animal taxonomic groups, we show that urban communities generally consist of smaller species. In addition, although we show urban warming for three habitat types and associated reduced community-weighted mean body sizes for four taxa, three taxa display a shift to larger species along the urbanization gradients. Our results show that the general trend towards smaller-sized species is overruled by filtering for larger species when there is positive covariation between size and dispersal, a process that can mitigate the low connectivity of ecological resources in urban settings<sup>5</sup>. We thus demonstrate that the urban-heat-island effect and urban habitat fragmentation are associated with contrasting community-level shifts in body size that critically depend on the association between body size and dispersal. Because body size determines the structure and dynamics of ecological networks<sup>1</sup>, such shifts may affect urban ecosystem function.**

Body size is a fundamental species trait relating to space use and key life-history features such as longevity and fecundity<sup>6</sup>. It also drives interspecific relationships, thus affecting ecological network dynamics<sup>1</sup>. Size-biased species loss has profound effects on ecosystem function<sup>7,8</sup>. Ectotherms rely on ambient conditions to achieve operational body temperatures<sup>9</sup>. Because higher ambient temperature increases metabolic rates and the associated costs for a given body size<sup>2</sup>, global climatic warming is expected to drive shifts to communities consisting of smaller species<sup>3</sup>.

Our planet is urbanizing quickly<sup>10</sup>, which is a primary example of human-induced rapid environmental change. Cities are urban heat islands characterized by increased temperatures that are decades ahead of global averages<sup>11</sup>. Not only are cities warmer than surrounding areas, but they also experience extensive fragmentation of (semi-)natural habitats, and both of these effects increase with percentage built-up cover (BUC; a proxy for urbanization)<sup>12,13</sup>. This provides an opportunity to study the opposing effects of size-dependent thermal tolerance and dispersal capacity, as larger body size favours dispersal in some, but not all, taxa.

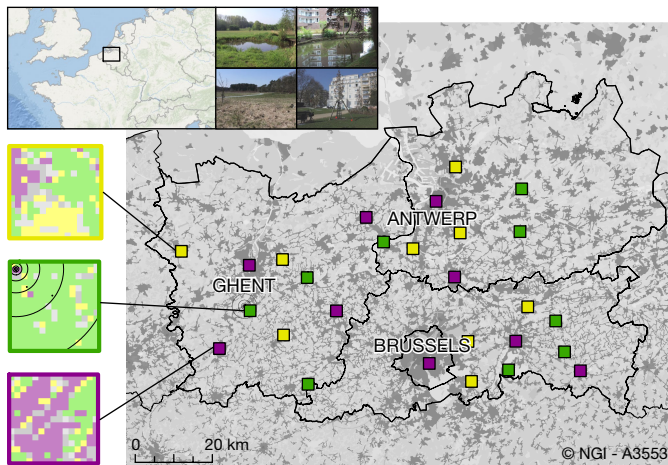
Here we test the hypothesis that urbanization causes shifts in community-level body size, and that these shifts are dictated by the community-specific association between body size and dispersal. We generally expect the urban-heat-island effect to drive shifts to species with smaller body sizes in communities of ectothermic species, in line with Atkinson's temperature-size rule<sup>14</sup>. For taxa characterized by a positive association between body size and dispersal, however, we also expect a filtering in favour of larger-bodied species associated with habitat fragmentation<sup>5,15</sup>. Filtering for increased mobility has been demonstrated for urban ground beetle and plant communities<sup>16,17</sup>. Hence, for taxa characterized by a positive body-size-dispersal link, we predict that the general community-level pattern of smaller species with increasing urbanization may be neutralized or even reversed.

To test our hypothesis, we engaged in an analysis of community-level shifts in body size across a broad range of both terrestrial and aquatic taxa along the same systematically sampled urbanization gradients. We studied the direction of change of community-level body size in ten taxa using a replicated, highly standardized and nested sampling design that covers urbanization gradients at seven spatial scales (50–3,200 m radii; Fig. 1). We sampled each taxon at up to 81 sites, sampling 95,001 individuals from 702 species, with species-specific body size varying by a factor of 400 (0.2–80 mm; Extended Data Table 1). Three of the ten groups are characterized by a positive association between body size and dispersal capacity (see Extended Data Table 1).

We show that the local temperature of pond, grassland and woodland habitats significantly increases with urbanization (linear mixed regression models,  $P < 0.002$ ; Extended Data Table 2). The intensity of these urban-heat-island effects is consistently larger during night and summer, in accordance with slower night-time city cooling and higher irradiation levels in summer<sup>18</sup> (Fig. 2, Extended Data Fig. 1, Extended Data Table 2). We also show that increased urbanization is linked to significant declines in habitat amount and the patch size of terrestrial habitats, and significant increases in distances among patches for both terrestrial and aquatic habitats (Pearson's  $r$  correlations,  $P \leq 0.020$ ; Extended Data Fig. 2).

Confirming our metabolism-based prediction that interspecific mean body size decreases with increasing temperature, urban communities for four out of the seven taxa (ground spiders, ground beetles, weevils and cladocerans) that did not have a positive size-dispersal link display reduced community-weighted mean body size (CWMBs). For ostracods, bdelloid rotifers and web spiders, no relationship with urbanization is found. By contrast, all three taxa with positive size-dispersal links display increased CWMBs in response to urbanization

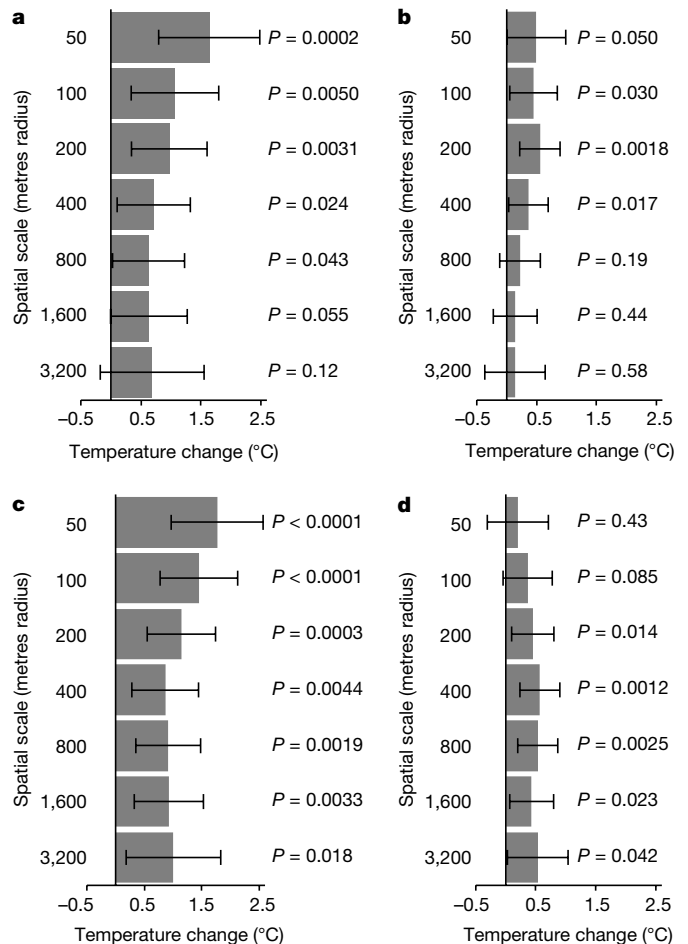
<sup>1</sup>Behavioural Ecology and Conservation Group, Biodiversity Research Centre, Earth and Life Institute, Université catholique de Louvain, Louvain-la-Neuve, Belgium. <sup>2</sup>Laboratory of Aquatic Ecology, Evolution and Conservation, KU Leuven, Leuven, Belgium. <sup>3</sup>Evolutionary Ecology Group, University of Antwerp, Antwerp, Belgium. <sup>4</sup>Directorate Taxonomy and Phylogeny, Royal Belgian Institute of Natural Sciences, Brussels, Belgium. <sup>5</sup>Terrestrial Ecology Unit, Biology Department, Ghent University, Ghent, Belgium. <sup>6</sup>Directorate Natural Environment, Royal Belgian Institute of Natural Sciences, Brussels, Belgium. <sup>7</sup>ECOBIO (Ecosystèmes, biodiversité, évolution), CNRS, Université de Rennes, Rennes, France. <sup>8</sup>Laboratory of Evolutionary Genetics and Ecology, URBE, NAXYS, University of Namur, Namur, Belgium. <sup>9</sup>National Research Council, Institute of Ecosystem Study, Verbania-Pallanza, Italy. <sup>10</sup>German Centre for Integrative Biodiversity Research (iDiv), Halle-Jena-Leipzig, Germany. <sup>11</sup>Helmholtz Centre for Environmental Research (UFZ), Department of Community Ecology, Halle, Germany. <sup>12</sup>Centre of Research in Limnology, Ichthyology and Aquaculture/PEA, State University of Maringá, Maringá, Brazil. <sup>13</sup>Limnology Research Unit, Biology Department, Ghent University, Ghent, Belgium. <sup>14</sup>Department of Life Sciences and Systems Biology, University of Turin, Turin, Italy. <sup>15</sup>Zoology Research Group, University of Hasselt, Hasselt, Belgium. <sup>16</sup>These authors jointly supervised this work: Luc De Meester, Hans Van Dyck. \*e-mail: th.merckx@gmail.com



**Fig. 1 | Map of the study area.** The configuration of 27 landscape-scale sampling plots (nine urban, magenta; nine semi-urban, yellow; nine non-urban, green) on an urbanization background (light and dark corresponds to non-urban and urban gradient, respectively). Solid lines refer to administrative province borders. Three plots are enlarged, showing the distribution of local subplot types within each plot. Subplots allowed sampling using a nested design that covers urbanization gradients at both the landscape and local scale. Different sets of subplots were selected among taxa, so that subplots always contained the corresponding habitats. Urbanization was quantified as the percentage BUC (assessed using a reference map with building contours; LRD, <https://www.agiv.be/international/en/products/grb-en>) for each sample site at seven spatial scales (50–3,200 m radii, depicted as black circles around one of the three sample sites of one non-urban plot). Photographs depict sites in a non-urban and urban subplot, for both aquatic (top row) and terrestrial (bottom row) systems. Corine Land Cover map for northern Belgium, copyright NGI (2017); map of western Europe, Ocean Basemap, copyright Esri.

(Figs. 3, 4, Extended Data Table 3). The positive shifts in size observed for these taxa are in line with our prediction that increased urbanization-mediated habitat fragmentation selects for larger species in taxa with positive size–dispersal links.

The Benjamini–Hochberg procedure<sup>19</sup>, which controls for false positives, confirms that all seven responses are significant at the study-wide level. Comparing the percentage changes in body size over a percentage BUC gradient of 0–25% shows a marked difference between taxa with a positive size–dispersal link ( $13.6\% \pm 8.3\%$  (mean  $\pm$  s.e.m.) body size increase) versus the other taxa ( $15.6\% \pm 5.3\%$  body size decrease) (weighted two-sided analysis of variance (ANOVA):  $F_{1,8} = 12.38$ ;  $P = 0.0079$ ). These community-level shifts in body size occur independently of shifts in species abundance and diversity along the urbanization gradients. For example, reduced diversity is apparent for taxa that display positive and negative size shifts, as well as for web spiders that lack a size shift. By contrast, cladocerans show size reduction without diversity change (Extended Data Table 4). For butterflies, macro-moths and orthopterans (that is, taxa with a positive size–dispersal link), the increase in the CWMBs ranges from 7% to 21% depending on the taxon, whereas size reductions of ground beetles, weevils and ground spiders (that is, terrestrial taxa with non-positive size–dispersal links) range from –18% to –21% over an urbanization gradient of 0–25% BUC (Fig. 3). The cladocerans display the largest size reduction (–44%), in accordance with the temperature–size response generally being stronger in aquatic species than in terrestrial species as a result of the greater oxygen limitation in water<sup>20</sup>. However, the size reduction for the ostracods is much smaller (–13%) and non-significant (linear mixed regression model,  $P = 0.10$ ), and for the rotifers no size shift is found. The absence of a size shift for the microscopic rotifers might indicate that their small size allows for sufficient oxygen exchange between warm, low-oxygen environments and body tissues, so that no community shift to smaller

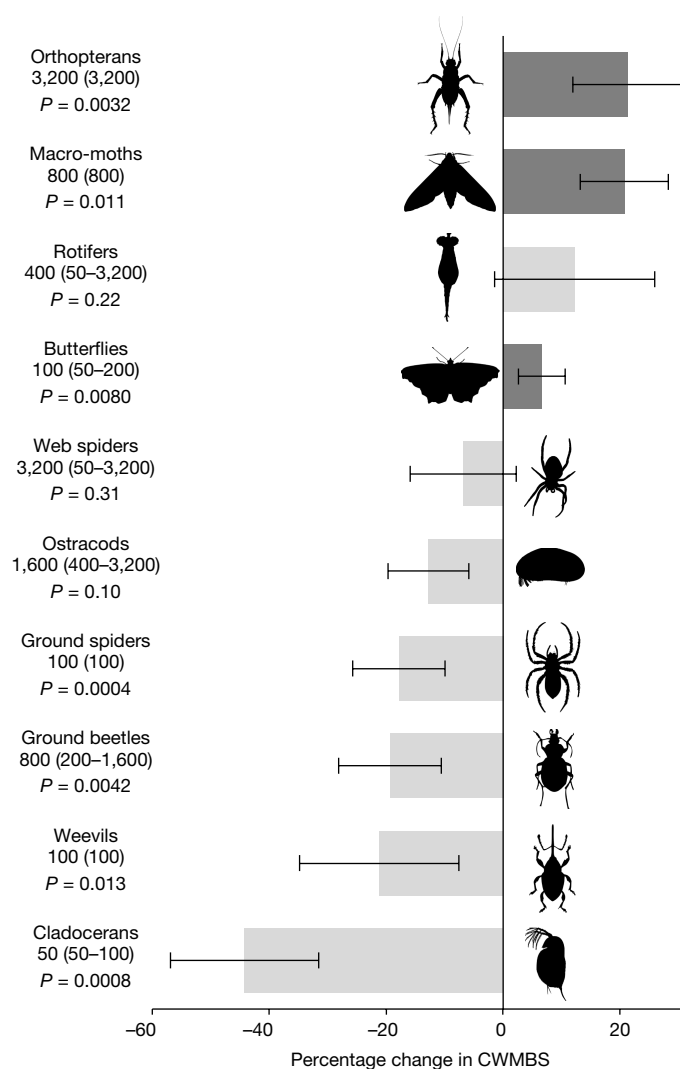


**Fig. 2 | Micro-climatic urban-heat-island effects.** a–d, Mean temperature increase (°C) when comparing sites that differed by 25% BUC. Effects for pond, grassland and woodland habitats are merged, and displayed separately for seven spatial scales (50–3,200 m radii at which urbanization was quantified). a, Summer diurnal ( $n = 103$  sites). b, Winter diurnal ( $n = 93$  sites). c, Summer nocturnal ( $n = 103$  sites). d, Winter nocturnal ( $n = 93$  sites). Temperature averages were analysed in relation to site-specific percentage BUC values using linear mixed regression models. We calculated mean changes in temperature over a 0–25% BUC gradient, on the basis of model-estimated intercepts and slopes. Error bars represent the range of temperature change on the basis of these slopes with 95% confidence intervals.

body sizes is induced by increased temperature. The absence of a size shift for web spiders may be explained by behavioural flexibility in their extended phenotype, as modified web designs help web-spider communities to adapt to urbanization-induced lower average body size of aerial dipteran prey<sup>21</sup>.

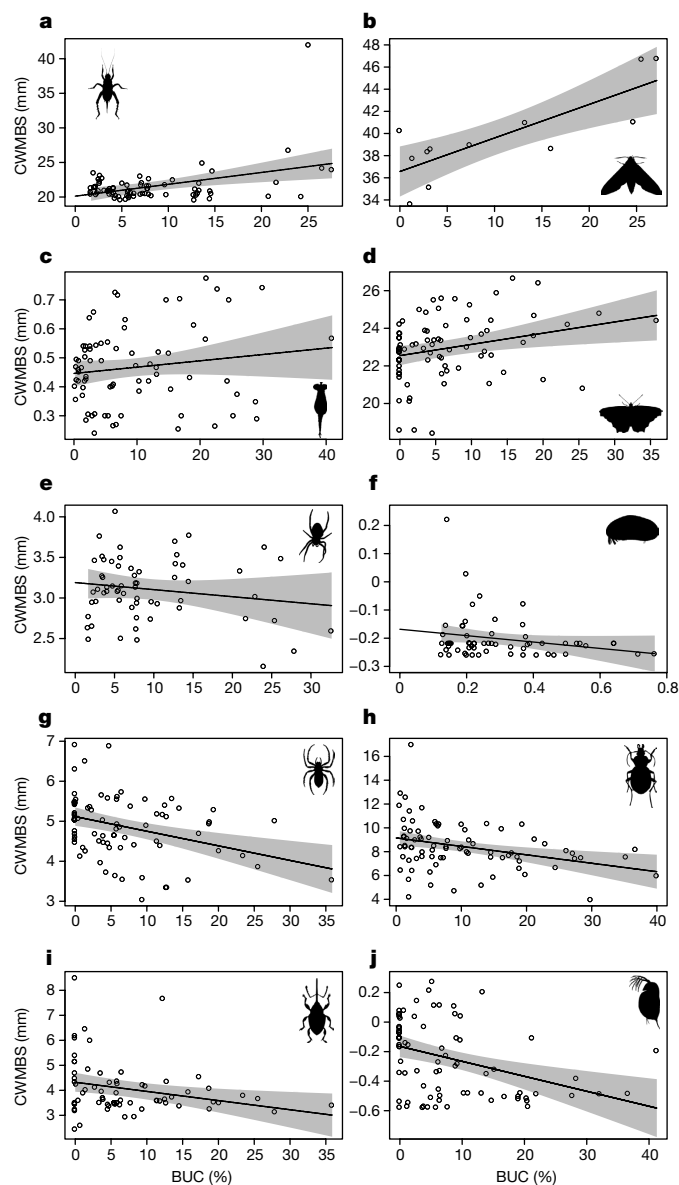
Our multi-scale approach allows the pinpointing of the spatial scales at which urbanization best explains the observed effects. During winter, the urban-heat-island effect fades with increasing spatial scale during the day but not at night, whereas during summer both diurnal and nocturnal urban-heat-island effects are more pronounced at small scales (Fig. 2, Extended Data Fig. 1, Extended Data Table 2). The spatial scale at which most of the variation in CWMBs is explained varied considerably among taxa, with effects for smaller-sized taxa prevailing at small spatial scales (Figs. 3, 4, Extended Data Table 3).

Urbanization induces biodiversity loss and biotic homogenization<sup>10</sup> (see also Extended Data Table 4). Here, we demonstrate that urbanization also leads to community-wide shifts in body size for the majority of studied species groups. The size reductions within aquatic and terrestrial taxa follow metabolic rules in line with the urban-heat-island effect, especially as our data on various pollutants suggest no correlation with percentage BUC (data not shown). By



**Fig. 3 | Taxon-specific percentage change in CWMBS for a 0–25% change in urbanization.** Modelled extent of the mean percentage change in CWMBS for each taxon when comparing sites that differed by 25% BUC. CWMBS was analysed for each taxon ( $n = 76, 12, 75, 80, 62, 60, 81, 81, 68$  and  $80$  biologically independent communities from top to bottom) in relation to site-specific percentage BUC values using linear mixed regression models. We calculated the percentage change in CWMBS over a 0–25% BUC gradient for each taxon, at the spatial scale of the best-fitting model, on basis of the modelled intercept and slope. Error bars are based on the 95% confidence intervals of the regression slopes. No adjustments were made for multiple comparisons, but the Benjamini–Hochberg procedure<sup>19</sup> confirmed that all significant responses were truly significant. Numbers indicate the scale (metres radius) of the best-fitting model, the range given between brackets indicates the radii of the models which are within the confidence set. Dark grey bars correspond to taxa with a positive size–dispersal link. The first nine animal silhouettes are from PhyloPic (<http://www.phylopic.org>) and fall under CC-BY 3.0 licences; credits (top to bottom): M. Broussard, G. Monger, D. Fontaneto et al., G. Monger, B. Lang, M. Dahirel, S. Harmer et al., T. Assmann et al., E. Schmidt & M. Dahirel; the bottom silhouette is in the public domain.

contrast, the increased fragmentation that is a result of urbanization appears to cause size increases for taxa with positive size–dispersal links. Hence, our multi-taxa study provides evidence of bi-directional shifts in community body size. In addition to the interspecific patterns reported here, shifts in body size can also occur at the intraspecific level, through phenotypic plasticity and genotypic change<sup>22–24</sup>. Our results should enable mechanistic studies that elucidate the cause of the variation in the observed shifts in body



**Fig. 4 | Taxon-specific plots of CWMBS as a function of urbanization.** a–j, Modelled CWMBS (mm) values of all taxa plotted against percentage BUC at the spatial scale (metres radius) of the best-fitting model. a, Orthopterans (spatial scale of best-fitting model: 3,200 m). b, Macro-moths (800 m). c, Rotifers (400 m). d, Butterflies (100 m). e, Web spiders (3,200 m). f, Ostracods (1,600 m). g, Ground spiders (100 m). h, Ground beetles (800 m). i, Weevils (100 m). j, Cladocerans (50 m).  $n = 76, 12, 75, 80, 62, 60, 81, 81, 68$  and  $80$  biologically independent communities for a–j, respectively. CWMBS values are log-transformed for ostracods and cladocerans (depicted range: 0.55–1.66 and 0.26–1.89 mm, respectively), and for the former, the percentage BUC values are also transformed (depicted range: 1.5–47.8%). Modelled linear regression slopes with 95% confidence intervals (shaded regions) are shown. Image credits as in Fig. 3.

size along urban gradients and quantify their functional effects in urban ecosystems. A better insight into the mechanisms behind shifts in body size will allow prediction of the intertwined effects of climate change and urbanization on the body-size distribution of communities.

The size-biased species loss reported here is expected to strongly affect ecosystem function<sup>7,8</sup>. If taxa in urban areas are represented by smaller or larger species, ecosystem structure and function will be affected in several ways. Metabolic theory and a recent artificial-selection experiment predict that shifted size distributions affect whole-ecosystem properties such as primary productivity, carbon cycling and

decomposition<sup>25,26</sup>. Shifts in body size also translate into altered life histories, demographic rates and interspecific relationships<sup>1,2</sup>. For example, consumer–resource dynamics have recently been modelled for warming-related intraspecific size shifts mediated by phenotypic plasticity<sup>27</sup>. A clear-cut effect of shifts in body size on ecosystem function can be predicted for cladoceran zooplankton. Smaller-sized cladoceran communities are typified by reduced densities of large *Daphnia* species (highly efficient filter feeders that consume phytoplankton), and are thus less able to maintain top-down control on algal blooms than larger-sized communities<sup>28</sup>. Also, the observed shifts in macro-moth body-size distributions may be functionally linked to flowering plant diversity through pollination<sup>29,30</sup>.

The shifts in body size that we observe across a range of animal taxa will be directly relevant to future efforts to understand, predict and mediate population resilience, trophic interactions, and ecosystem function in urban ecosystems<sup>31,32</sup>. Such insights will be essential to design the biodiverse towns and cities of the future. For example, urban planners could mitigate the micro-climatic effects and habitat fragmentation that result from urbanization with measures implemented at multiple spatial scales. Such interventions could involve the creation and/or modification of urban ponds and urban green infrastructure to increase the amount and quality of habitats<sup>33</sup>. Doing so would reduce the urban-heat-island effect and favour dispersal, and hence gene flow, in urban animal populations. Our results indicate that such impacts would maintain variation in the body-size distributions of urban communities and potentially mitigate the effect that shifts in body size may have on ecosystem function.

## Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0140-0>.

Received: 11 July 2017; Accepted: 11 April 2018;

Published online: 23 May 2018

- Woodward, G. et al. Body size in ecological networks. *Trends Ecol. Evol.* **20**, 402–409 (2005).
- Brown, J. H., Gillooly, J. F., Allen, A. P., Savage, V. M. & West, G. B. Toward a metabolic theory of ecology. *Ecology* **85**, 1771–1789 (2004).
- Scheffers, B. R. et al. The broad footprint of climate change from genes to biomes to people. *Science* **354**, aaf7671 (2016).
- Alberti, M., Marzluff, J. & Hunt, V. M. Urban driven phenotypic changes: empirical observations and theoretical implications for eco-evolutionary feedback. *Phil. Trans. R. Soc. Lond. B* **372**, 20160029 (2017).
- Cheptou, P. O., Hargreaves, A. L., Bonte, D. & Jacquemyn, H. Adaptation to fragmentation: evolutionary dynamics driven by human influences. *Phil. Trans. R. Soc. Lond. B* **372**, 20160037 (2017).
- Chown, S. L. & Gaston, K. J. Body size variation in insects: a macroecological perspective. *Biol. Rev. Camb. Philos. Soc.* **85**, 139–169 (2010).
- Kalinkat, G. et al. Body masses, functional responses and predator–prey stability. *Ecol. Lett.* **16**, 1126–1134 (2013).
- Brose, U. et al. Predicting the consequences of species loss using size-structured biodiversity approaches. *Biol. Rev. Camb. Philos. Soc.* **92**, 684–697 (2017).
- Horne, C. R., Hirst, A. G. & Atkinson, D. Seasonal body size reductions with warming covary with major body size gradients in arthropod species. *Proc. R. Soc. B* **284**, 20170238 (2017).
- Seto, K. C., Güneralp, B. & Hutyra, L. R. Global forecasts of urban expansion to 2030 and direct impacts on biodiversity and carbon pools. *Proc. Natl Acad. Sci. USA* **109**, 16083–16088 (2012).
- Youngsteadt, E., Dale, A. G., Terando, A. J., Dunn, R. R. & Frank, S. D. Do cities simulate climate change? A comparison of herbivore response to urban and global warming. *Glob. Chang. Biol.* **21**, 97–105 (2015).
- Ward, K., Lauf, S., Kleinschmit, B. & Endlicher, W. Heat waves and urban heat islands in Europe: a review of relevant drivers. *Sci. Total Environ.* **569–570**, 527–539 (2016).
- Niemelä, J. *Urban Ecology: Patterns, Processes, and Applications* (Oxford Univ. Press, Oxford, 2011).

- Atkinson, D. Temperature and organism size: a biological law for ectotherms? *Adv. Ecol. Res.* **25**, 1–58 (1994).
- Bonte, D. & Doherty, M. Dispersal: a central and independent trait in life history. *Oikos* **126**, 472–479 (2017).
- Piano, E. et al. Urbanization drives community shifts towards thermophilic and dispersive species at local and landscape scales. *Glob. Chang. Biol.* **23**, 2554–2564 (2017).
- Concepción, E. D., Moretti, M., Altermatt, F., Nobis, M. P. & Obrist, M. K. Impacts of urbanisation on biodiversity: the role of species mobility, degree of specialisation and spatial scale. *Oikos* **124**, 1571–1582 (2015).
- Arnfield, A. J. Two decades of urban climate research: a review of turbulence, exchanges of energy and water, and the urban heat island. *Int. J. Climatol.* **23**, 1–26 (2003).
- Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B Stat. Methodol.* **57**, 289–300 (1995).
- Forster, J., Hirst, A. G. & Atkinson, D. Warming-induced reductions in body size are greater in aquatic than terrestrial species. *Proc. Natl Acad. Sci. USA* **109**, 19310–19314 (2012).
- Dahirel, M., Dierick, J., De Cock, M. & Bonte, D. Intraspecific variation shapes community-level behavioral responses to urbanization in spiders. *Ecology* **98**, 2379–2390 (2017).
- McDonnell, M. J. & Hahs, A. K. Adaptation and adaptedness of organisms to urban environments. *Annu. Rev. Ecol. Syst.* **46**, 261–280 (2015).
- Alberti, M. et al. Global urban signatures of phenotypic change in animal and plant populations. *Proc. Natl Acad. Sci. USA* **114**, 8951–8956 (2017).
- Johnson, M. T. J. & Munshi-South, J. Evolution of life in urban environments. *Science* **358**, eaam8327 (2017).
- Schramski, J. R., Dell, A. I., Grady, J. M., Sibily, R. M. & Brown, J. H. Metabolic theory predicts whole-ecosystem properties. *Proc. Natl Acad. Sci. USA* **112**, 2617–2622 (2015).
- Malerba, M. E., White, C. R. & Marshall, D. J. Eco-energetic consequences of evolutionary shifts in body size. *Ecol. Lett.* **21**, 54–62 (2018).
- Osmond, M. M. et al. Warming-induced changes to body size stabilize consumer–resource dynamics. *Am. Nat.* **189**, 718–725 (2017).
- Gianuca, A. T., Pantel, J. H. & De Meester, L. Disentangling the effect of body size and phylogenetic distances on zooplankton top-down control of algae. *Proc. R. Soc. B* **283**, 20160487 (2016).
- Agosta, S. J. & Janzen, D. H. Body size distributions of large Costa Rican dry forest moths and the underlying relationship between plant and pollinator morphology. *Oikos* **108**, 183–193 (2005).
- Biesmeijer, J. C. et al. Parallel declines in pollinators and insect-pollinated plants in Britain and the Netherlands. *Science* **313**, 351–354 (2006).
- Duffy, J. E. Biodiversity loss, trophic skew and ecosystem functioning. *Ecol. Lett.* **6**, 680–687 (2003).
- Webb, C. T., Hoeting, J. A., Ames, G. M., Pyne, M. I. & LeRoy Poff, N. A structured and dynamic framework to advance traits-based theory and prediction in ecology. *Ecol. Lett.* **13**, 267–283 (2010).
- Threlfall, C. G. et al. Increasing biodiversity in urban green spaces through simple vegetation interventions. *J. Appl. Ecol.* **54**, 1874–1883 (2017).

**Acknowledgements** We thank M. De Cock, J. Dierick, P. Limbourg, E. van den Berg, M. Van Kerckvoorde and P. Vantighem for help with sampling and identification of species. This is publication BRC419 of the Biodiversity Research Centre (UCL/ELI). This research is part of the SPEEDY-project, funded by the Interuniversity Attraction Poles program of the Belgian Science Policy Office BELSPO (IAP-grant P7/04).

**Reviewer information** Nature thanks M. McDonnell and the other anonymous reviewer(s) for their contribution to the peer review of this work.

**Author contributions** T.M., L.D.M. and H.V.D. conceived the study. C.S. and L.D.M. coordinated the consortium. T.M., C.S., A.K., L.F.B., T.B., D.B., K.I.B., M.C., M.D., N.D., K.D.W., J.M.T.E., D.F., A.T.G., L.G., F.H., J.H., L.L., K.M., E.M., E.P., R.S., I.S. and K.V.D. contributed to sampling and data collection. T.M. and A.K. performed the analyses. H.M. selected study plots, calculated fragmentation variables and designed the study area map. T.M. wrote the first draft of the manuscript with all authors contributing substantially to revisions.

**Competing interests** The authors declare no competing interests.

## Additional information

**Extended data** is available for this paper at <https://doi.org/10.1038/s41586-018-0140-0>.

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41586-018-0140-0>.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>.

**Correspondence and requests for materials** should be addressed to T.M.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## METHODS

**Sampling design.** Sampling was performed according to a nested design in which a local urbanization gradient (three classes: non-urban, semi-urban and urban) was repeatedly sampled within landscapes distributed along a landscape-scale urbanization gradient (three classes: non-urban, semi-urban and urban). For each of ten taxa a total of up to 81 local-scale subplots ( $200 \times 200 \text{ m}^2$ ) were sampled within 27 landscape-scale plots ( $3 \times 3 \text{ km}^2$ ) situated in an 8,140  $\text{km}^2$  study area in northern Belgium (Fig. 1, Extended Data Table 1). The average human population density of this highly urbanized area amounts to 693 individuals per  $\text{km}^2$ , with cities and urban sprawl embedded within an agricultural and semi-natural matrix<sup>34</sup>. As a proxy for urbanization we used percentage BUC, which was assessed in a geographic information system (GIS) using an object-oriented reference map of Flanders with the precise contours of all buildings, excluding roads and parking infrastructures, as a vectorial layer. Given that only buildings are considered, 15% BUC can be considered highly urbanized. Within each of the nine urban ( $\text{BUC} > 15\%$ ), nine semi-urban ( $5\% < \text{BUC} < 10\%$ ) and nine non-urban ( $\text{BUC} < 3\%$ ) plots, one urban, one semi-urban and one non-urban subplot were chosen using identical BUC cut-off values, for a total of 81 subplots. Within each subplot, and for each of the ten taxa, a single grassland, woodland or pond habitat patch was targeted for sampling during the most appropriate season for each taxon (Extended Data Table 1). As each taxon was sampled in only one of three habitat types (that is, grassland, woodland or ponds), it was often impossible to sample all taxa within the same  $200 \times 200 \text{ m}^2$  subplots. As such, independent subplots containing the corresponding habitats were sometimes selected among taxa, but these subplots were always of the same urbanization level and located within the same  $3 \times 3 \text{ km}^2$  plot.

The classification of urban, semi-urban and non-urban (sub)plots on the basis of BUC cut-off values was used to establish the nested sampling design, which allowed samples to display a wide range of urbanization values at both local (subplot) and landscape (plot) scales. To increase precision in the data analysis, however, we moved away from having BUC as a class variable with three levels, and instead quantified BUC as a continuous variable, at seven spatial scales around the sampling site (50, 100, 200, 400, 800, 1,600 and 3,200 m radii). Owing to our nested design, BUC values at small scales were not correlated with values at large scales, hence allowing the pinpointing of the scales at which the effects of urbanization are most pronounced.

Using this highly replicated, nested sampling design, our sampling effort involved counting and assigning 95,001 individuals to 702 species in ten taxa: (i) aquatic: cladocerans and ostracods sampled in pond habitats; (ii) limno-terrestrial: aquatic bdelloid rotifers sampled within the water layers of terrestrial *Xanthoria* lichens; and (iii) terrestrial: butterflies, orthopterans (that is, grasshoppers and bush crickets), macro-moths, ground beetles, weevils, web spiders and ground spiders sampled in grassland and woodland habitats (Extended Data Table 1). No statistical methods were used to predetermine sample size and the investigators were not blinded to allocation during experiments and outcome assessment.

**Urban-heat-island effect.** The urban-heat-island effect was quantified using hourly temperature readings that were collected automatically across 104 sampling sites for the three habitat types in which the ten taxa were sampled: ponds, grasslands and woodlands. Aquatic probes (HOB0, TidbiT v2 UTBI-001; HOBOWare ONSET; resolution:  $0.02^\circ\text{C}$ ) were attached to a floating device to log temperatures at 15 cm depth for 18 ponds (27 November 2014–29 November 2015). Terrestrial probes (iButton, Thermochron DS1923, Maxim Integrated; resolution:  $0.06^\circ\text{C}$ ) logged air temperature at 20 cm height near 59 pitfall sites (that is, grassland habitat; 8 May 2014–20 September 2015; 59 and 49 sites during summer and winter, respectively) and 27 macro-moth sampling sites (that is, woodland habitat; 1 April 2015–20 March 2016; 26 sites each during summer and winter). For each day, temperature averages of twelve diurnal (07:00–18:00) and twelve nocturnal (19:00–06:00) readings were calculated, which were labelled as summer from 21 March–20 September, and as winter from 21 September–20 March.

**Habitat fragmentation.** Correlations between urbanization (BUC) and three habitat fragmentation variables (that is, habitat coverage, mean size of habitat patches, and mean nearest-neighbour distance among habitat patches) were quantified using Pearson's  $r$  coefficients (Extended Data Fig. 2). This was done at a  $3 \times 3 \text{ km}^2$  plot scale, on the basis of detailed land-use data from all 27 sampling plots (Fig. 1), and separately for terrestrial (that is, all types of (semi-)natural habitat) and aquatic habitat (that is, all pond types)<sup>35,36</sup>. Eutrophied, mono-specific intensive grasslands as well as orchards, plantations and conifer woodlands were not retained for analyses.

**Statistical analyses.** Temperature averages were analysed in relation to site-specific urbanization (BUC) values and habitat type (grassland, woodland and pond) using linear mixed regression models (R package lme4). We ran separate models for both seasons (summer and winter) and for both day and night conditions (diurnal and nocturnal). Site ID and date (nested within year) were included as random factors. We used a multi-scale approach, running separate models with BUC values quantified at seven spatial scales (50–3,200 m radii).  $P$  values for the

fixed effects were obtained using likelihood-ratio tests of nested models that were fitted with maximum-likelihood and parameter estimates from restricted maximum-likelihood models. Residual plots were always visually inspected to evaluate the fit of models, and we compared maximum-likelihood-based AICc values (R package AICcmodavg) to select a confidence set of models for which the AICc values did not differ substantially from the value of the best-fitting model, using  $\Delta\text{AICc} \leq 2$  as a criterion<sup>37</sup>.

CWMS was calculated for a given site as the average of the species-specific body sizes (mm) for all locally sampled species, weighted by species abundance. The raw data for calculating this metric are species-level count data for all taxa in all sites (based on taxon-specific sampling and identification protocols) and mean species-specific body-size values extracted from the literature or, in the case of web spiders and cladocerans, from our own measurements (Extended Data Table 1). An increase in CWMS with increasing urbanization implies that the species assemblage of the site is increasingly composed of individuals belonging to larger species along the gradient from communities in more rural sites to communities in more urban sites. Our CWMS index hence reflects the relative composition of large versus small species in local communities, and we use it here to quantify community response to urbanization. Although every sampling method introduces some bias in relative species abundances, the extent of the bias should be similar for non-urban and urban sampling sites. Therefore, using the relative species abundances that we obtained via sampling to calculate the CWMS is appropriate to look into the relative effects of urbanization.

CWMS was analysed for each taxon in relation to site-specific urbanization (BUC) values using linear mixed regression models with restricted maximum-likelihood estimation (R package lme4). Plot ID was used as a random variable to account for potential spatial autocorrelation of variables among sites belonging to the same landscape-scale plot. CWMS values were  $\log_{10}$ -transformed for cladocerans and ostracods. For ostracods, we also transformed BUC values by taking the arcsine of square-rooted BUC values, which resulted in residual plots with a more homogeneous distribution. Analyses for the other taxa were run with untransformed data as residual plots proved to be homogeneous. The residual plots for orthopterans, ostracods and ground beetles each displayed one outlying data point, and the residual plot for weevils displayed two such points. Because these five data points are legitimate (that is, they are not due to measurement, data or sampling errors) we assessed their effect on the consistency of the regressions in the model output. Filtering these data points out of the regressions showed (i) that the best-fitting models remained linked to the identical spatial scales, (ii) that the positive slope for orthopterans remained positive and the negative slopes for the other taxa remained negative, and (iii) that the significance levels stayed equal for ground beetles and ostracods, got stronger for weevils, and decreased but remained significant for orthopterans. Because those five data points are legitimate and do not have a qualitative effect on the output, we opted to retain them in the analyses. We used a multi-scale approach, running separate models with BUC values quantified at seven spatial scales (50–3,200 m radii). For each taxon, we then selected the model (and hence the spatial scale) that fitted the data best using maximum-likelihood-based AICc values (R package AICcmodavg). We also retained a confidence set of the models for which the AICc values did not differ substantially from the value of the best model using  $\Delta\text{AICc} \leq 2$  as a criterion<sup>37</sup>.

For each taxon, and at the spatial scale of the best-fitting model, we calculated the percentage change (with 95% confidence interval) in CWMS over a 0–25% BUC gradient, on the basis of the modelled intercept and slope, or of back-transformed values for ostracods and cladocerans (Fig. 3). These values were then contrasted for taxa with a positive size–dispersal link against all other taxa using two-sided ANOVA, with the inverse of the taxon-specific error bars as weights to account for the differences among taxa in the variance of the estimated percentage change. This weighted ANOVA allows testing of the percentage change values for taxa with a positive size–dispersal link to determine whether they are significantly different from those of all other taxa.

All statistical analyses were performed using R v.3.2.3<sup>38</sup>.

**Reporting summary.** Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

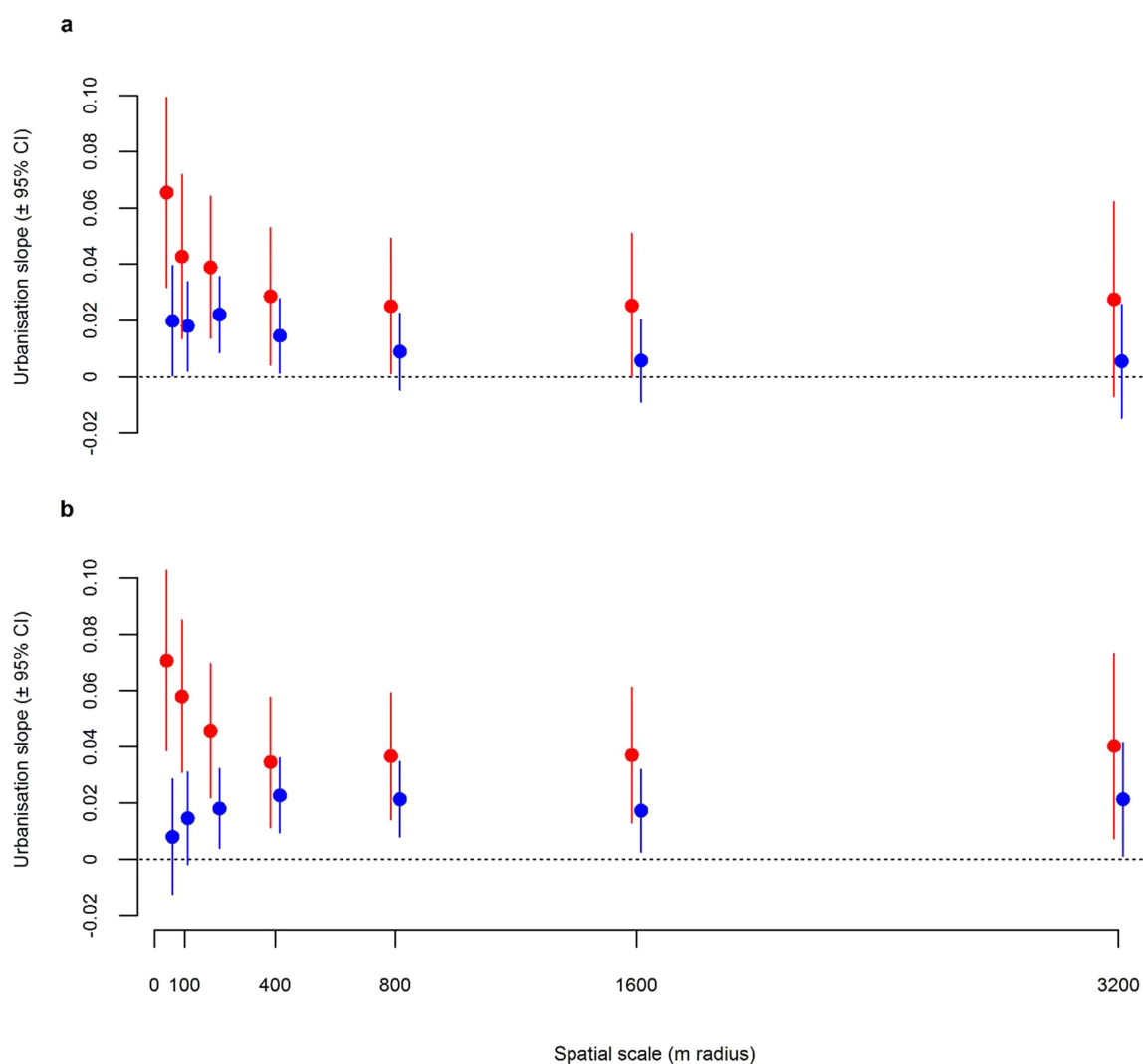
**Data availability.** The datasets generated and analysed during the current study are available from the corresponding author upon reasonable request.

34. IBZ. *Bevolkingscijfers per Provincie en per Gemeente op 1 Januari 2017* [http://www.ibz.rn.fgov.be/fileadmin/user\\_upload/fr/pop/statistiques/population-bevolking-20170101.pdf](http://www.ibz.rn.fgov.be/fileadmin/user_upload/fr/pop/statistiques/population-bevolking-20170101.pdf) (Federale Overheidsdienst Binnenlandse Zaken, 2017).

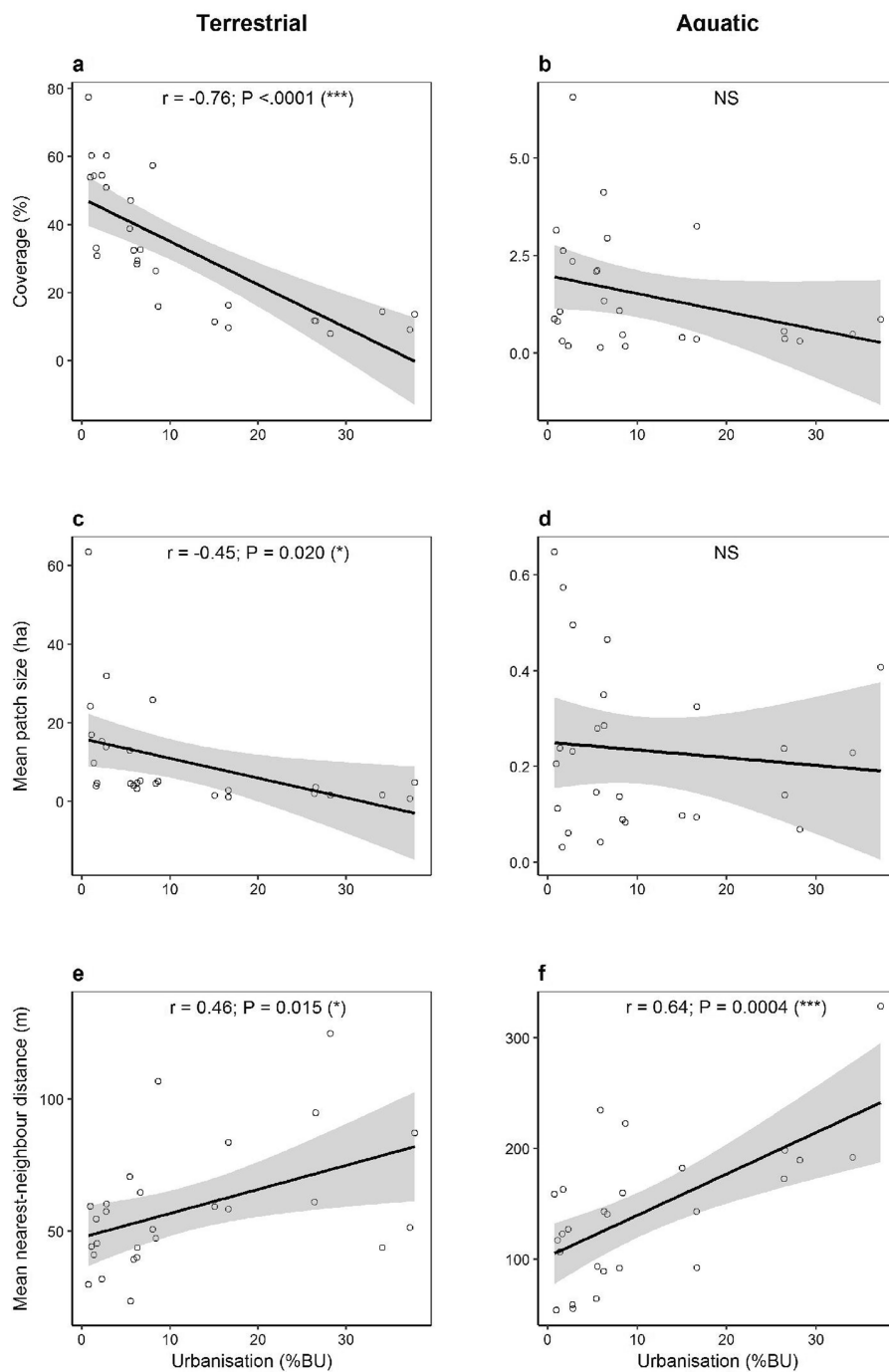
35. De Saeger, S. et al. *Biologische Waarderingskaart en Natura 2000 Habitatkaart* (Instituut voor Natuur en Bosonderzoek, Brussels, 2016).

36. Packet, J. et al. *Watervlakken Versie 1.0: Polygonenkaart van Stilstaand Water in Vlaanderen. Inhoud en Metadata van een Nieuw Instrument voor Water-, Milieu- en Natuurbeleid* (Instituut voor Natuur en Bosonderzoek, Brussels, 2017).

37. Burnham, K. P. & Anderson, D. R. *Model Selection and Multimodel Inference: a Practical Information-Theoretic Approach* (Springer Science & Business Media, New York, 2003).
38. R Development Core Team. *R: A Language and Environment for Statistical Computing* <https://www.r-project.org/> (R Foundation for Statistical Computing, Vienna, 2015).
39. Bink, F. A. *Ecologische Atlas van de Dagvlinders van Noordwest-Europa* (Schuyt & Co, Haarlem, 1992).
40. Sekar, S. A meta-analysis of the traits affecting dispersal ability in butterflies: can wingspan be used as a proxy? *J. Anim. Ecol.* **81**, 174–184 (2012).
41. Stevens, V. M., Trochet, A., Van Dyck, H., Clobert, J. & Baguette, M. How is dispersal integrated in life histories: a quantitative analysis using butterflies. *Ecol. Lett.* **15**, 74–86 (2012).
42. Stevens, V. M. et al. A comparative analysis of dispersal syndromes in terrestrial and semi-terrestrial animals. *Ecol. Lett.* **17**, 1039–1052 (2014).
43. Manley, C. *British Moths and Butterflies: a Photographic Guide* (A & C Black, London, 2008).
44. Nieminen, M., Rita, H. & Uuvana, P. Body size and migration rate in moths. *Ecography* **22**, 697–707 (1999).
45. Slade, E. M. et al. Life-history traits and landscape characteristics predict macro-moth responses to forest fragmentation. *Ecology* **94**, 1519–1530 (2013).
46. Reinhardt, K., Köhler, G., Maas, S. & Detzel, P. Low dispersal ability and habitat specificity promote extinctions in rare but not in widespread species: the Orthoptera of Germany. *Ecography* **28**, 593–602 (2005).
47. Roberts, M. J. *The Spiders of Great Britain and Ireland: Compact Edition* (Apollo Books, Vester Skerninge, 2009).
48. Turin, H. *De Nederlandse Loopkevers: Verspreiding en Ecologie* (KNNV, Zeist, 2000).
49. Duff, A. G. et al. *Beetles of Britain and Ireland 4* (A. G. Duff, West Runton, 2016).
50. Donner, J. *Ordnung Bdelloidea. Bestimmungsbücher zur Bodenfauna Europas* (Akademie-Verlag, Berlin, 1965).
51. Fontaneto, D. *Biogeography of Microscopic Organisms: is Everything Small Everywhere?* (Cambridge Univ. Press, Cambridge, 2011).
52. Meisch, C. *Freshwater Ostracoda of Western and Central Europe* (Spektrum Akademischer, Heidelberg, 2000).
53. Bilton, D. T., Freeland, J. R. & Okamura, B. Dispersal in freshwater invertebrates. *Annu. Rev. Ecol. Syst.* **32**, 159–181 (2001).
54. De Bie, T. et al. Body size and dispersal mode as key traits determining metacommunity structure of aquatic organisms. *Ecol. Lett.* **15**, 740–747 (2012).
55. Brans, K. I. et al. Eco-evolutionary dynamics in urbanized landscapes: evolution, species sorting and the change in zooplankton body size along urbanization gradients. *Phil. Trans. R. Soc. Lond. B* **372**, 20160030 (2017).
56. Gianuca, A. T. et al. Taxonomic, functional and phylogenetic metacommunity ecology of cladoceran zooplankton along urbanization gradients. *Ecography* **41**, 183–194 (2018).



**Extended Data Fig. 1 | Micro-climatic urban-heat-island effect strengths. a, b,** Slopes of the urban-heat-island effects (measured as the increase in temperature ( $^{\circ}\text{C}$ ) per 1% increase in percentage BUC) as a function of spatial scale (the radius at which urbanization was quantified) with 95% confidence intervals (CI). Separate measurements are shown for summer (red) and winter (blue) using merged readings for pond, grassland and woodland habitats ( $n = 104$  sites). **a**, Diurnal measurements. **b**, Nocturnal measurements. Data points are offset from one another horizontally to improve clarity.



**Extended Data Fig. 2 | Correlations between urbanization and habitat fragmentation.** Correlations between urbanization (measured as the percentage BUC) and three habitat fragmentation variables: habitat coverage (**a**, **b**), mean size of habitat patches (**c**, **d**), and mean nearest-neighbour distance among habitat patches (**e**, **f**). Separate plots are shown

for terrestrial (that is, all types of (semi-)natural habitat, **a**, **c**, **e**) and aquatic (that is, all pond types, **b**, **d**, **f**) habitats ( $n = 27$  landscape-scale sampling plots). Pearson's  $r$  coefficients and  $P$  values are indicated; not significant (NS),  $P > 0.1$ ; \* $P < 0.05$ ; \*\*\* $P < 0.001$ .

Extended Data Table 1 | Taxon-specific details of sampling procedures, body-size data and size–dispersal links

Taxon	Sampling method	Sites	N	S	Body size (proxy) (mm)	Body size data	Size-dispersal link
Butterflies	Visual counts in grassland while walking 20 minutes ('Pollard walk')/subplot, with occasional netting and visual inspections; each site sampled three times during July till early September 2014; up to 18 sites/day (simultaneous sampling by two researchers)	81	4,413	23	Forewing length	Means <sup>39</sup> , also for sexually or seasonally dimorphic species	Positive <sup>40–42</sup>
Macro-moths	Full-night light-trapping with one Heath trap (6W) at woodland; each site sampled 11 times during August till early September 2014 and during April, July and August 2015; four sites simultaneously/night; identification of within-trap samples during early mornings down to species-level, except for <i>Hoplodrina</i> and <i>Mesapamea</i> sp.	12	3,067	202	Wing span	Means <sup>43</sup> (+ <a href="http://www.lepidoptera.eu">www.lepidoptera.eu</a> ), also for sexually dimorphic species; but males only for three species with flightless females	Positive <sup>42,44,45</sup>
Orthopterans	Auditive counts in grassland of male grasshoppers and bush crickets while walking 20 minutes/subplot, with occasional visual inspections; each site sampled three times during July till early September 2014; up to 18 sites/day (simultaneous sampling by two researchers)	81	10,302	8	Body length	Means (without wings nor appendages) <sup>46</sup>	Positive <sup>46</sup> (our subset)
Web spiders	Visual and complete exploration of subplots to collect and store every individual in 70% ethanol until identification via a microscope of all adult specimens; three sites sampled/day during September 2014	62	2,456	18	Cephalothorax width	Means of all captured adult spiders; microscope-measured	Neutral (bell-shaped) <sup>42</sup>
Ground spiders	Pitfall trapping, simultaneously at all sites with two pitfalls/site placed in grassy, open habitats from April till August 2013. Identification via microscope of all adult specimens, stored in 70% ethanol	81	27,763	184	Body length	Values of females <sup>47</sup> (+ <a href="http://www.araneae.unibe.ch">www.araneae.unibe.ch</a> )	Neutral (bell-shaped) <sup>42</sup>
Ground beetles	Identical to ground spider sampling	81	7,604	128	Body length	Means <sup>48</sup>	Neutral <sup>42</sup>
Weevils	Identical to ground spider sampling	78	2,600	73	Body length	Means of minimum + maximum values <sup>49</sup>	Neutral
Rotifers	Community sampling of bdelloid rotifers recovered from dormancy four hours after hydration of one <i>Xanthoria</i> lichen thalli of 2.5 cm <sup>2</sup> in a petri dish, a period known to recover all dormant individuals; each site sampled once during July 2013; up to 18 sites/day	81	4,936	21	Body length	Maximum recorded lengths in literature; mostly from original species descriptions and <sup>50</sup>	Neutral <sup>51</sup>
Ostracods	Handnet sampling in one pond/subplot at up to nine ponds/day from mid-August till mid-September 2014. Individual ostracods were sorted from the bulk sample under a microscope to a minimum of 50 individuals, in cases where ostracods were present. Rarefaction analyses showed that the samples were representative for the ostracod communities	81	3,111	17	Body length	Values of females <sup>52</sup>	Negative or neutral <sup>53,54</sup>
Cladocerans	Zooplankton sampling with tube sampler in one pond/subplot, collecting 5 l water at each of eight locations/pond, integrating the entire water column from close to bottom till surface; crustacean zooplankton for density assessment is filtered through a 64 µm conical net; samples are collected in 60 ml vials and fixed with formalin (2 ml in 48 ml of sample); min. 300 individuals were identified/sample ( <i>Daphnia longispina</i> , <i>D. galeata</i> and <i>D. hyalina</i> were combined in the <i>D. longispina</i> complex); individual counts were volume-corrected for each sample; 15 random individuals/species occurring in each sample were measured (if less individuals present/species, all were measured). Sampling was conducted from 29 May till 10 July 2013. Three ponds (one plot) were sampled/day, with plot sampling randomized over the sampling period. Detailed information <sup>55,56</sup>	81	28,749	28	Body length	Means (eye top till tail spine base) of up to 15 individuals per species per sample, with <i>Ceriodaphnia</i> values combined into one category. Means from all ponds were averaged further	Negative or neutral <sup>53,54</sup>

N, counted individuals; S, cumulative species richness.

Some of the sampling methods and data on body size and size–dispersal links are from previously published work<sup>39–56</sup>.

Extended Data Table 2 | Model output of average temperature in relation to urbanization and habitat type

	Scale	$\Delta AICc$	Fixed effect	$\chi^2$ test	P value	Estimate $\pm$ s.e.m.
<b>Diurnal summer</b>	50	0.00	BUC x Habitat	$\chi^2_2 = 4.05$	$P = 0.13$	$0.0655 \pm 0.0172$
			BUC	$\chi^2_1 = 13.96$	$P = 0.0001$	
			Habitat	$\chi^2_2 = 39.67$	$P < 0.0001$	
<b>Nocturnal summer</b>	50	0.00	BUC x Habitat	$\chi^2_2 = 2.73$	$P = 0.25$	$0.0706 \pm 0.0163$
			BUC	$\chi^2_1 = 17.66$	$P < 0.0001$	
			Habitat	$\chi^2_2 = 82.37$	$P < 0.0001$	
	100	0.91	BUC x Habitat	$\chi^2_2 = 0.14$	$P = 0.93$	$0.0579 \pm 0.0138$
			BUC	$\chi^2_1 = 16.76$	$P < 0.0001$	
			Habitat	$\chi^2_2 = 83.19$	$P < 0.0001$	
<b>Diurnal winter</b>	200	0.00	BUC x Habitat	$\chi^2_2 = 0.21$	$P = 0.89$	$0.0221 \pm 0.0069$
			BUC	$\chi^2_1 = 10.15$	$P = 0.001$	
			Habitat	$\chi^2_2 = 5.45$	$P = 0.06$	
<b>Nocturnal winter</b>	400	0.00	BUC x Habitat	$\chi^2_2 = 0.21$	$P = 0.89$	$0.0227 \pm 0.0068$
			BUC	$\chi^2_1 = 10.94$	$P = 0.0009$	
			Habitat	$\chi^2_2 = 76.57$	$P < 0.0001$	
	800	1.33	BUC x Habitat	$\chi^2_2 = 0.39$	$P = 0.82$	$0.0213 \pm 0.0068$
			BUC	$\chi^2_1 = 9.61$	$P = 0.0019$	
			Habitat	$\chi^2_2 = 77.55$	$P < 0.0001$	

Output of linear mixed regression models that test the relationship between average local ambient temperatures and the interaction between percentage BUC and habitat type (pond, grassland and woodland) ( $n = 104$  sites). Only the output for the confidence set of models ( $\Delta AICc \leq 2$ ) is given, with scale referring to the associated radius scale (in metres) of percentage BUC. Model estimates ( $\pm$ s.e.m.) for percentage BUC regression coefficients are provided. Model output consistently shows clear temperature differences among habitats and a clear positive effect of urbanization on temperature, irrespective of habitat type (as shown by the non-significant interactions).

Extended Data Table 3 | Model output of CWMBS in relation to urbanization

Taxon	Scale	$\Delta AICc$	F test	P value	Intercept	Slope
Orthopterans	50	9.24	$F_{1,61.0} = 0.14$	$P = 0.71$	$21.458 \pm 0.436$	$0.015 \pm 0.040$
	100	8.91	$F_{1,53.8} = 0.48$	$P = 0.49$	$21.370 \pm 0.457$	$0.024 \pm 0.035$
	200	6.83	$F_{1,59.9} = 2.49$	$P = 0.12$	$21.104 \pm 0.476$	$0.052 \pm 0.033$
	400	8.29	$F_{1,73.9} = 1.01$	$P = 0.32$	$21.203 \pm 0.502$	$0.035 \pm 0.034$
	800	7.85	$F_{1,55.0} = 1.43$	$P = 0.24$	$21.084 \pm 0.527$	$0.047 \pm 0.038$
	1,600	5.67	$F_{1,34.1} = 3.64$	$P = 0.065$	$20.752 \pm 0.548$	$0.081 \pm 0.042$
	<b>3,200</b>	<b>0.00</b>	<b><math>F_{1,26.7} = 10.46</math></b>	<b><math>P = 0.0032</math></b>	<b><math>20.122 \pm 0.549</math></b>	<b><math>0.171 \pm 0.053</math></b>
Macro-moths	50	10.14	$F_{1,8.6} = 1.72$	$P = 0.22$	$38.883 \pm 1.293$	$0.227 \pm 0.148$
	100	7.70	$F_{1,7.0} = 4.26$	$P = 0.078$	$37.949 \pm 1.336$	$0.282 \pm 0.123$
	200	3.98	$F_{1,9.1} = 7.64$	$P = 0.022$	$36.721 \pm 1.245$	$0.335 \pm 0.103$
	400	3.04	$F_{1,7.4} = 8.96$	$P = 0.019$	$36.886 \pm 1.140$	$0.306 \pm 0.087$
	<b>800</b>	<b>0.00</b>	<b><math>F_{1,4.6} = 16.84</math></b>	<b><math>P = 0.011</math></b>	<b><math>36.566 \pm 1.016</math></b>	<b><math>0.303 \pm 0.070</math></b>
	1,600	2.78	$F_{1,4.1} = 12.41$	$P = 0.023$	$36.889 \pm 1.122$	$0.273 \pm 0.076$
	<b>3,200</b>	<b>4.57</b>	<b><math>F_{1,4.0} = 9.57</math></b>	<b><math>P = 0.036</math></b>	<b><math>36.996 \pm 1.227</math></b>	<b><math>0.319 \pm 0.103</math></b>
Rotifers	50	1.08	$F_{1,70.9} = 0.55$	$P = 0.46$	$0.456 \pm 0.021$	$0.002 \pm 0.002$
	100	1.24	$F_{1,67.5} = 0.41$	$P = 0.52$	$0.457 \pm 0.022$	$0.001 \pm 0.002$
	200	0.23	$F_{1,72.1} = 1.35$	$P = 0.25$	$0.447 \pm 0.023$	$0.002 \pm 0.002$
	<b>400</b>	<b>0.00</b>	<b><math>F_{1,53.9} = 1.55</math></b>	<b><math>P = 0.22</math></b>	<b><math>0.446 \pm 0.022</math></b>	<b><math>0.002 \pm 0.002</math></b>
	800	0.69	$F_{1,37.2} = 0.92$	$P = 0.34$	$0.450 \pm 0.023$	$0.002 \pm 0.002$
	1,600	0.36	$F_{1,27.5} = 1.26$	$P = 0.27$	$0.447 \pm 0.024$	$0.002 \pm 0.002$
	<b>3,200</b>	<b>0.81</b>	<b><math>F_{1,23.2} = 0.84</math></b>	<b><math>P = 0.37</math></b>	<b><math>0.449 \pm 0.025</math></b>	<b><math>0.002 \pm 0.002</math></b>
Butterflies	50	0.41	$F_{1,71.2} = 7.25$	$P = 0.0088$	$22.637 \pm 0.230$	$0.068 \pm 0.025$
	<b>100</b>	<b>0.00</b>	<b><math>F_{1,59.6} = 7.53</math></b>	<b><math>P = 0.0080</math></b>	<b><math>22.546 \pm 0.257</math></b>	<b><math>0.060 \pm 0.022</math></b>
	200	1.49	$F_{1,66.2} = 6.27$	$P = 0.015$	$22.511 \pm 0.285$	$0.050 \pm 0.020$
	400	6.50	$F_{1,77.5} = 1.10$	$P = 0.30$	$22.720 \pm 0.308$	$0.022 \pm 0.020$
	800	7.45	$F_{1,53.8} = 0.05$	$P = 0.82$	$22.885 \pm 0.318$	$0.005 \pm 0.022$
	1,600	7.17	$F_{1,33.4} = 0.27$	$P = 0.61$	$23.063 \pm 0.332$	$-0.013 \pm 0.024$
	<b>3,200</b>	<b>7.39</b>	<b><math>F_{1,27.7} = 0.07</math></b>	<b><math>P = 0.80</math></b>	<b><math>23.007 \pm 0.360</math></b>	<b><math>-0.008 \pm 0.032</math></b>
Web spiders	50	0.14	$F_{1,46.5} = 0.96$	$P = 0.33$	$3.133 \pm 0.067$	$-0.003 \pm 0.003$
	100	0.83	$F_{1,42.7} = 0.32$	$P = 0.58$	$3.128 \pm 0.073$	$-0.002 \pm 0.004$
	200	0.83	$F_{1,47.2} = 0.31$	$P = 0.58$	$3.131 \pm 0.076$	$-0.003 \pm 0.005$
	400	0.76	$F_{1,59.6} = 0.35$	$P = 0.56$	$3.138 \pm 0.081$	$-0.003 \pm 0.005$
	800	0.93	$F_{1,48.7} = 0.19$	$P = 0.66$	$3.134 \pm 0.087$	$-0.002 \pm 0.005$
	1,600	1.08	$F_{1,28.7} = 0.05$	$P = 0.82$	$3.124 \pm 0.093$	$-0.002 \pm 0.006$
	<b>3,200</b>	<b>0.00</b>	<b><math>F_{1,22.3} = 1.06</math></b>	<b><math>P = 0.31</math></b>	<b><math>3.190 \pm 0.101</math></b>	<b><math>-0.009 \pm 0.008</math></b>
Ostracods	50	3.03	$F_{1,57.0} = 0.00$	$P = 0.97$	$-0.202 \pm 0.016$	$-0.002 \pm 0.062$
	100	2.72	$F_{1,57.2} = 0.27$	$P = 0.60$	$-0.194 \pm 0.019$	$-0.031 \pm 0.058$
	200	2.09	$F_{1,58.0} = 0.88$	$P = 0.35$	$-0.186 \pm 0.020$	$-0.053 \pm 0.055$
	400	1.19	$F_{1,51.4} = 1.71$	$P = 0.20$	$-0.179 \pm 0.020$	$-0.075 \pm 0.055$
	800	0.89	$F_{1,43.5} = 1.98$	$P = 0.17$	$-0.176 \pm 0.021$	$-0.084 \pm 0.058$
	<b>1,600</b>	<b>0.00</b>	<b><math>F_{1,32.2} = 2.83</math></b>	<b><math>P = 0.10</math></b>	<b><math>-0.168 \pm 0.022</math></b>	<b><math>-0.113 \pm 0.066</math></b>
	<b>3,200</b>	<b>0.22</b>	<b><math>F_{1,26.8} = 2.67</math></b>	<b><math>P = 0.11</math></b>	<b><math>-0.160 \pm 0.027</math></b>	<b><math>-0.148 \pm 0.089</math></b>
Ground spiders	50	7.95	$F_{1,69.6} = 5.07$	$P = 0.028$	$4.993 \pm 0.113$	$-0.027 \pm 0.012$
	<b>100</b>	<b>0.00</b>	<b><math>F_{1,60.0} = 13.82</math></b>	<b><math>P = 0.0004</math></b>	<b><math>5.116 \pm 0.116</math></b>	<b><math>-0.036 \pm 0.010</math></b>
	200	2.81	$F_{1,71.6} = 9.83$	$P = 0.0025$	$5.127 \pm 0.122$	$-0.029 \pm 0.009$
	400	5.30	$F_{1,72.5} = 6.92$	$P = 0.010$	$5.113 \pm 0.127$	$-0.024 \pm 0.009$
	800	5.40	$F_{1,46.3} = 6.86$	$P = 0.012$	$5.123 \pm 0.130$	$-0.025 \pm 0.009$
	1,600	6.80	$F_{1,31.2} = 5.57$	$P = 0.025$	$5.110 \pm 0.134$	$-0.023 \pm 0.010$
	<b>3,200</b>	<b>7.23</b>	<b><math>F_{1,26.9} = 5.34</math></b>	<b><math>P = 0.029</math></b>	<b><math>5.124 \pm 0.139</math></b>	<b><math>-0.029 \pm 0.013</math></b>
Ground beetles	50	7.17	$F_{1,75.2} = 2.02$	$P = 0.16$	$8.646 \pm 0.276$	$-0.047 \pm 0.032$
	100	3.66	$F_{1,64.4} = 5.57$	$P = 0.021$	$8.894 \pm 0.294$	$-0.069 \pm 0.029$
	200	1.79	$F_{1,76.5} = 7.39$	$P = 0.0081$	$9.045 \pm 0.310$	$-0.071 \pm 0.025$
	400	1.58	$F_{1,65.8} = 7.44$	$P = 0.0082$	$9.080 \pm 0.319$	$-0.066 \pm 0.023$
	<b>800</b>	<b>0.00</b>	<b><math>F_{1,41.4} = 9.19</math></b>	<b><math>P = 0.0042</math></b>	<b><math>9.152 \pm 0.318</math></b>	<b><math>-0.071 \pm 0.023</math></b>
	1,600	1.39	$F_{1,29.7} = 7.94$	$P = 0.0085$	$9.124 \pm 0.326$	$-0.068 \pm 0.024$
	<b>3,200</b>	<b>5.51</b>	<b><math>F_{1,26.7} = 3.70</math></b>	<b><math>P = 0.065</math></b>	<b><math>8.976 \pm 0.360</math></b>	<b><math>-0.063 \pm 0.033</math></b>
Weevils	50	4.20	$F_{1,55.7} = 2.02$	$P = 0.16$	$4.170 \pm 0.178$	$-0.024 \pm 0.017$
	<b>100</b>	<b>0.00</b>	<b><math>F_{1,48.1} = 6.59</math></b>	<b><math>P = 0.013</math></b>	<b><math>4.323 \pm 0.190</math></b>	<b><math>-0.037 \pm 0.014</math></b>
	200	2.21	$F_{1,56.5} = 3.92$	$P = 0.053$	$4.309 \pm 0.200$	$-0.028 \pm 0.014$
	400	2.37	$F_{1,66.0} = 3.71$	$P = 0.059$	$4.330 \pm 0.210$	$-0.027 \pm 0.013$
	800	3.65	$F_{1,50.8} = 2.51$	$P = 0.12$	$4.308 \pm 0.223$	$-0.024 \pm 0.014$
	1,600	5.20	$F_{1,32.2} = 0.99$	$P = 0.33$	$4.230 \pm 0.236$	$-0.017 \pm 0.016$
	<b>3,200</b>	<b>5.68</b>	<b><math>F_{1,26.5} = 0.51</math></b>	<b><math>P = 0.48</math></b>	<b><math>4.193 \pm 0.251</math></b>	<b><math>-0.016 \pm 0.022</math></b>
Cladocerans	<b>50</b>	<b>0.00</b>	<b><math>F_{1,78.7} = 12.37</math></b>	<b><math>P = 0.0008</math></b>	<b><math>-0.164 \pm 0.037</math></b>	<b><math>-0.010 \pm 0.003</math></b>
	100	0.32	$F_{1,78.9} = 12.72$	$P = 0.0006$	$-0.141 \pm 0.042$	$-0.009 \pm 0.003$
	200	4.48	$F_{1,73.6} = 8.74$	$P = 0.0042$	$-0.156 \pm 0.044$	$-0.007 \pm 0.002$
	400	5.35	$F_{1,63.8} = 7.82$	$P = 0.0068$	$-0.160 \pm 0.044$	$-0.007 \pm 0.002$
	800	8.69	$F_{1,45.1} = 3.67$	$P = 0.062$	$-0.186 \pm 0.044$	$-0.005 \pm 0.003$
	1,600	10.37	$F_{1,31.1} = 1.68$	$P = 0.20$	$-0.204 \pm 0.044$	$-0.004 \pm 0.003$
	<b>3,200</b>	<b>10.73</b>	<b><math>F_{1,26.6} = 1.19</math></b>	<b><math>P = 0.29</math></b>	<b><math>-0.204 \pm 0.048</math></b>	<b><math>-0.005 \pm 0.004</math></b>

Output of linear mixed regression models testing the relationship between CWMBS and percentage BUC at multiple scales (radii in metres), for ten taxa ( $n = 76, 12, 75, 80, 62, 60, 81, 81, 68$  and 80 biologically independent communities, top to bottom). Confidence sets of models ( $\Delta AICc \leq 2$ ) have grey shading and the best-fitting model output is given in bold. Modelled intercepts and slopes ( $\pm$ s.e.m.) are given.

Extended Data Table 4 | Model output of abundance and diversity measures in relation to urbanization

Taxon	<i>N/S/H</i>	Scale	<i>F</i> test	<i>P</i> value	% change 0–25% BUC
Orthopterans	t <i>N</i>	200	$F_{1,66.2} = 20.58$	$P < 0.0001$	-82.9
	<i>S</i>	400	$F_{1,78.5} = 16.24$	$P = 0.0001$	-34.5
	<i>H</i>	400	$F_{1,63.0} = 0.68$	$P = 0.41$	-8.4
Macro-moths	t <i>N</i>	3,200	$F_{1,4.0} = 52.6$	$P = 0.0019$	-89.2
	t <i>S</i>	3,200	$F_{1,4.0} = 108.1$	$P = 0.0005$	-82.7
	<i>H</i>	800	$F_{1,5.2} = 55.8$	$P = 0.0006$	-43.5
Rotifers	t <i>N</i>	400	$F_{1,67.1} = 2.1$	$P = 0.15$	+108.1
	t <i>S</i>	400	$F_{1,67.9} = 0.4$	$P = 0.53$	+15.6
	<i>H</i>	3,200	$F_{1,37.9} = 1.2$	$P = 0.28$	+38.8
Butterflies	t <i>N</i>	200	$F_{1,71.4} = 42.1$	$P < 0.0001$	-85.3
	<i>S</i>	200	$F_{1,69.7} = 54.2$	$P < 0.0001$	-59.1
	<i>H</i>	200	$F_{1,75.8} = 7.3$	$P = 0.0085$	-22.5
Web spiders	t <i>N</i>	200	$F_{1,54.1} = 7.9$	$P = 0.0069$	-18.3
	t <i>S</i>	200	$F_{1,53.8} = 15.1$	$P = 0.0003$	-29.2
	t <i>H</i>	200	$F_{1,54.0} = 12.3$	$P = 0.0009$	-21.1
Ostracods	t <i>N</i>	50	$F_{1,71.1} = 3.6$	$P = 0.06$	-69.2
	t <i>S</i>	50	$F_{1,71.3} = 2.1$	$P = 0.15$	-38.6
	<i>H</i>	1,600	$F_{1,35.1} = 2.2$	$P = 0.15$	-41.2
Ground spiders	<i>N</i>	100	$F_{1,65.1} = 5.7$	$P = 0.020$	-43.6
	<i>S</i>	800	$F_{1,43.3} = 2.3$	$P = 0.14$	-13.4
	<i>H</i>	3,200	$F_{1,26.6} = 12.3$	$P = 0.0016$	-20.3
Ground beetles	t <i>N</i>	800	$F_{1,47.2} = 5.8$	$P = 0.020$	-50.7
	t <i>S</i>	800	$F_{1,44.3} = 11.9$	$P = 0.0013$	-39.9
	t <i>H</i>	200	$F_{1,76.9} = 11.5$	$P = 0.0011$	-21.9
Weevils	t <i>N</i>	100	$F_{1,56.9} = 12.0$	$P = 0.0010$	+547.9
	t <i>S</i>	100	$F_{1,56.8} = 4.5$	$P = 0.038$	+99.2
	t <i>H</i>	400	$F_{1,63.9} = 0.7$	$P = 0.40$	+25.0
Cladocerans	t <i>N</i>	3,200	$F_{1,27.2} = 1.2$	$P = 0.29$	-68.5
	<i>S</i>	200	$F_{1,62.8} = 1.1$	$P = 0.29$	+12.7
	<i>H</i>	3,200	$F_{1,26.2} = 0.2$	$P = 0.65$	-11.4

Output of linear mixed regression models for ten taxa ( $n = 76, 12, 75, 80, 62, 60, 81, 81, 68$  and  $80$  biologically independent communities, top to bottom), testing the relationship between abundance (*N*) and two diversity measures (species richness (*S*) and Shannon index (*H*)) and percentage BUC at the spatial scale (radius in metres) providing the best-fitting models. *t* indicates that  $\log(x + 1)$  transformations improved residual fits. Modelled (back-transformed) percentage change across a 0–25% BUC gradient is also given.

# Pairwise and higher-order genetic interactions during the evolution of a tRNA

Júlia Domingo<sup>1</sup>, Guillaume Diss<sup>1</sup> & Ben Lehner<sup>1,2,3\*</sup>

**A central question in genetics and evolution is the extent to which the outcomes of mutations change depending on the genetic context in which they occur<sup>1–3</sup>. Pairwise interactions between mutations have been systematically mapped within<sup>4–18</sup> and between<sup>19</sup> genes, and have been shown to contribute substantially to phenotypic variation among individuals<sup>20</sup>. However, the extent to which genetic interactions themselves are stable or dynamic across genotypes is unclear<sup>21,22</sup>. Here we quantify more than 45,000 genetic interactions between the same 87 pairs of mutations across more than 500 closely related genotypes of a yeast tRNA. Notably, all pairs of mutations interacted in at least 9% of genetic backgrounds and all pairs switched from interacting positively to interacting negatively in different genotypes (false discovery rate < 0.1). Higher-order interactions are also abundant and dynamic across genotypes. The epistasis in this tRNA means that all individual mutations switch from detrimental to beneficial, even in closely related genotypes. As a consequence, accurate genetic prediction requires mutation effects to be measured across different genetic backgrounds and the use of higher-order epistatic terms.**

Genetic (epistatic) interactions have been extensively mapped between pairs of mutations within individual genes<sup>4–18</sup>, and also between individual alleles of many different genes<sup>19</sup>. However, the pairwise mapping of interactions only provides a limited view of genotype space, which has a vast combinatorial size<sup>22</sup>. Interactions between genes have been reported as only poorly or moderately conserved between species<sup>21</sup>. Moreover, analyses of the effects of combinations of mutations within individual genes have pointed to the importance of higher-order epistasis<sup>22–25</sup>, in which mutations interact beyond pairwise interactions to determine mutation effect.

To directly test the extent to which the effects of mutations and the interactions between mutations are stable or change depending upon the genotype in which they occur, we designed an experiment in which mutation effects and interactions are quantified across a large number of closely related genetic backgrounds. As a model system, we used the single-copy arginine-CCU tRNA (tRNA-Arg(CCU)) gene that is conditionally required for the growth of budding yeast (Extended Data Fig. 1a) and for which pairwise interactions have been previously mapped in one genetic background<sup>15</sup>. The small size of the gene allowed us to design a library that covered all 5,184 ( $2^6 \times 3^4$ ) genotypes containing the 14 nucleotide substitutions observed in ten positions in post-whole-genome-duplication yeast species<sup>26</sup> (Fig. 1a, b). Each genotype therefore varies from zero to a maximum of ten nucleotides divergence from the *Saccharomyces cerevisiae* tRNA sequence (Extended Data Fig. 1b). After transformation of the library into *S. cerevisiae*, we performed six selection experiments in parallel to quantify the relative fitness of each of the 5,184 variants under restrictive conditions (high temperature and 1 M NaCl) (Fig. 1c). The fitness of each genotype was quantified as the change in its abundance in each culture between the beginning and end of the competition period determined using deep sequencing with a hierarchical error model and normalized in log scale to the fitness of the *S. cerevisiae*

genotype (henceforth ‘fitness’). After filtering, we obtained fitness measurements for 4,176 variants (Supplementary Table 1) that correlated well across replicates (Fig. 1d). The median fitness declines as the number of mutations increases but there are still many combinations of mutations with high fitness amongst genotypes that are far from the reference genotype (Fig. 1e).

We first examined the fitness consequences of single mutations and how these change across different genetic backgrounds (Fig. 2a). In the *S. cerevisiae* genotype, six of the 14 individual mutations were detrimental (Fig. 2b). However, when the same 14 mutations were made in the tRNA genotypes of the other six extant species (these alternative ‘wild-type’ tRNAs have fitness very close to the *S. cerevisiae* tRNA when expressed in *S. cerevisiae*, Supplementary Table 2), their effects changed substantially (Fig. 2b). For example, the mutation C66A had no effect in the *S. cerevisiae* background but became detrimental in the *Candida glabrata* tRNA, which only differs by two substitutions (paired *t*-test,  $q = 0.006$ ,  $n = 6$ ). Indeed, 11 out of 14 mutations had effects that changed across these seven tRNAs from different species (Extended Data Fig. 2a, false discovery rate (FDR) < 0.1).

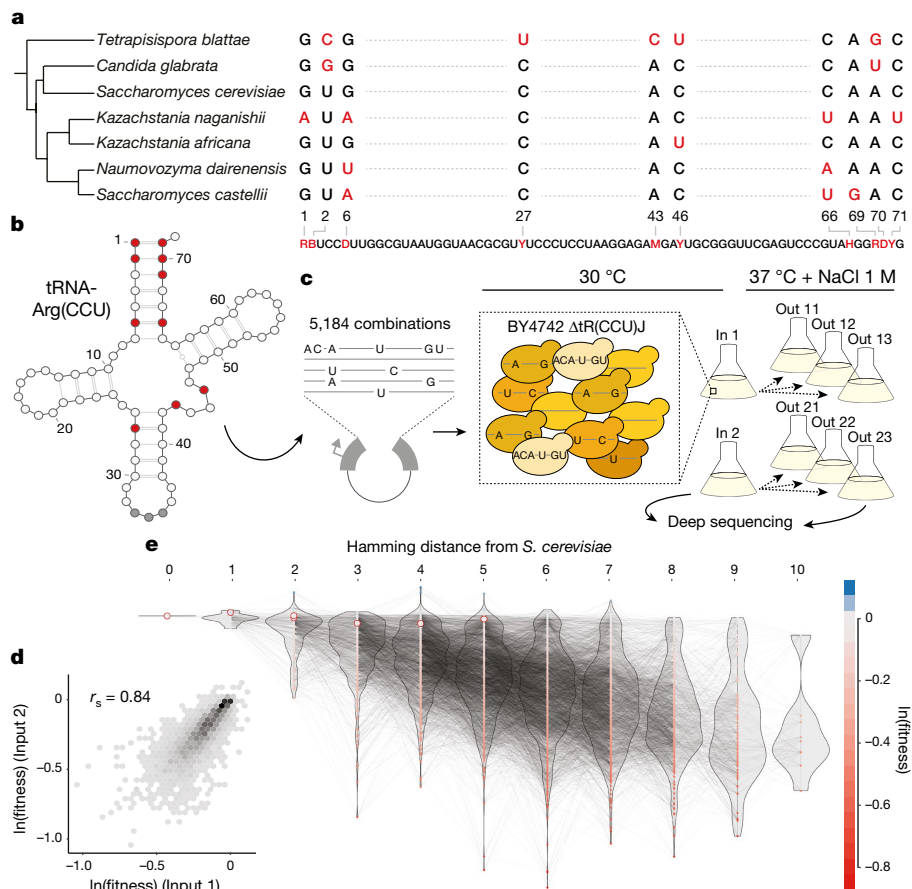
We next compared the effects of the single mutations across the complete set of genetic backgrounds in the library. In total, we tested each mutation in a median of 1,449 genetic backgrounds (minimum = 1,088, maximum = 1,993, Extended Data Fig. 1c, d). Notably, we found that every mutation was both detrimental and beneficial in a substantial number of genetic backgrounds (Fig. 2b, c, median number of backgrounds in which the less frequent sign was observed = 6.4%; minimum = 3.4%; maximum = 11.9% across all 14 mutations, FDR < 0.1,  $n = 21,450$ ). Restricting the analyses to background genotypes with high or intermediate fitness, to genotypes with high input read counts, or to genotypes with few mutations did not change this conclusion (Extended Data Fig. 2b). Thus, all mutations have effects that switch from beneficial to detrimental in closely related genotypes.

To investigate the interactions between mutations that underlie these changes in mutation effects, we first quantified pairwise genetic interactions between the 14 mutations, which is a total of 87 pairs in any genotype. We define epistasis as the difference between the fitness of each double mutant and the sum of the fitness of the two corresponding individual mutations. Consistent with previous results<sup>15</sup>, in the *S. cerevisiae* genotype, many pairs of mutations (40.2%, 35 out of 87) had combined fitness effects that were more detrimental than expected (negative epistasis) and only a few had effects that were less detrimental than expected (positive epistasis, 5.7%, 5 out of 87, FDR < 0.1, Fig. 3a). However, these interactions changed when they were tested in tRNAs from different species (Fig. 3b, c, Extended Data Fig. 3), with 83 out of the 87 interactions differing across the species ( $n = 1,000$  paired *t*-tests, FDR < 0.1, Extended Data Fig. 4).

We next analysed how the 87 interactions changed across all the genetic backgrounds in the library. Each interaction was quantified in a median of 506 genetic backgrounds (minimum = 240, maximum = 946, Extended Data Fig. 1d). Notably, all 87 interactions switched from positive to negative in a substantial proportion of the

<sup>1</sup>Systems Biology Program, Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Barcelona, Spain. <sup>2</sup>Universitat Pompeu Fabra (UPF), Barcelona, Spain.

<sup>3</sup>Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain. \*e-mail: ben.lehner@crgeu



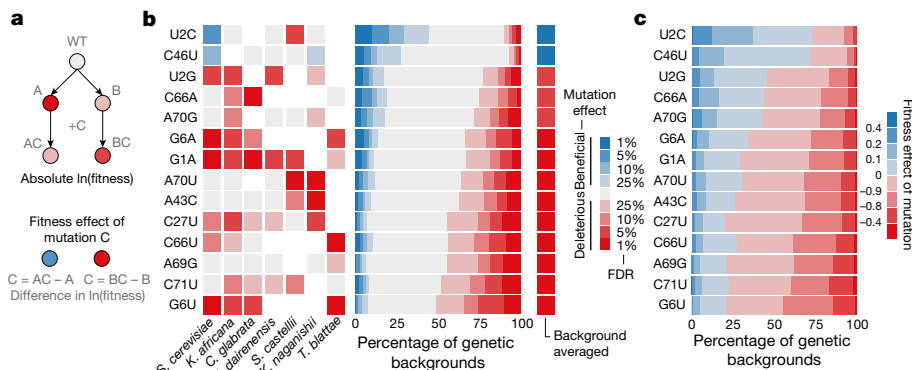
**Fig. 1 | Combinatorially complete fitness landscape of a tRNA.**

**a**, Species phylogenetic tree<sup>26</sup> and multiple sequence alignment of the tRNA-Arg(CCU) orthologues. Variable positions across the seven yeast species with the synthesized library are shown below: R (A or G); B (C, G or T); D (A, G or T); Y (C or T); M (A or C); H (A, C or T). **b**, Secondary structure of *S. cerevisiae* tRNA-Arg(CCU) (varied positions indicated in red). **c**, Selection experiment and structure of the replicates. From each independent yeast transformation (input) three independent

genetic backgrounds (Fig. 3a). Restricting our analyses to genetic backgrounds with high or intermediate fitness, to combinations with high expected fitness or to genotypes with high input read counts did not change this conclusion (Extended Data Fig. 5b). Across all genetic backgrounds, positive and negative interactions were similarly prevalent (11.4% and 10.3% for positive and negative epistasis respectively, FDR < 0.1,  $n = 47,649$ ).

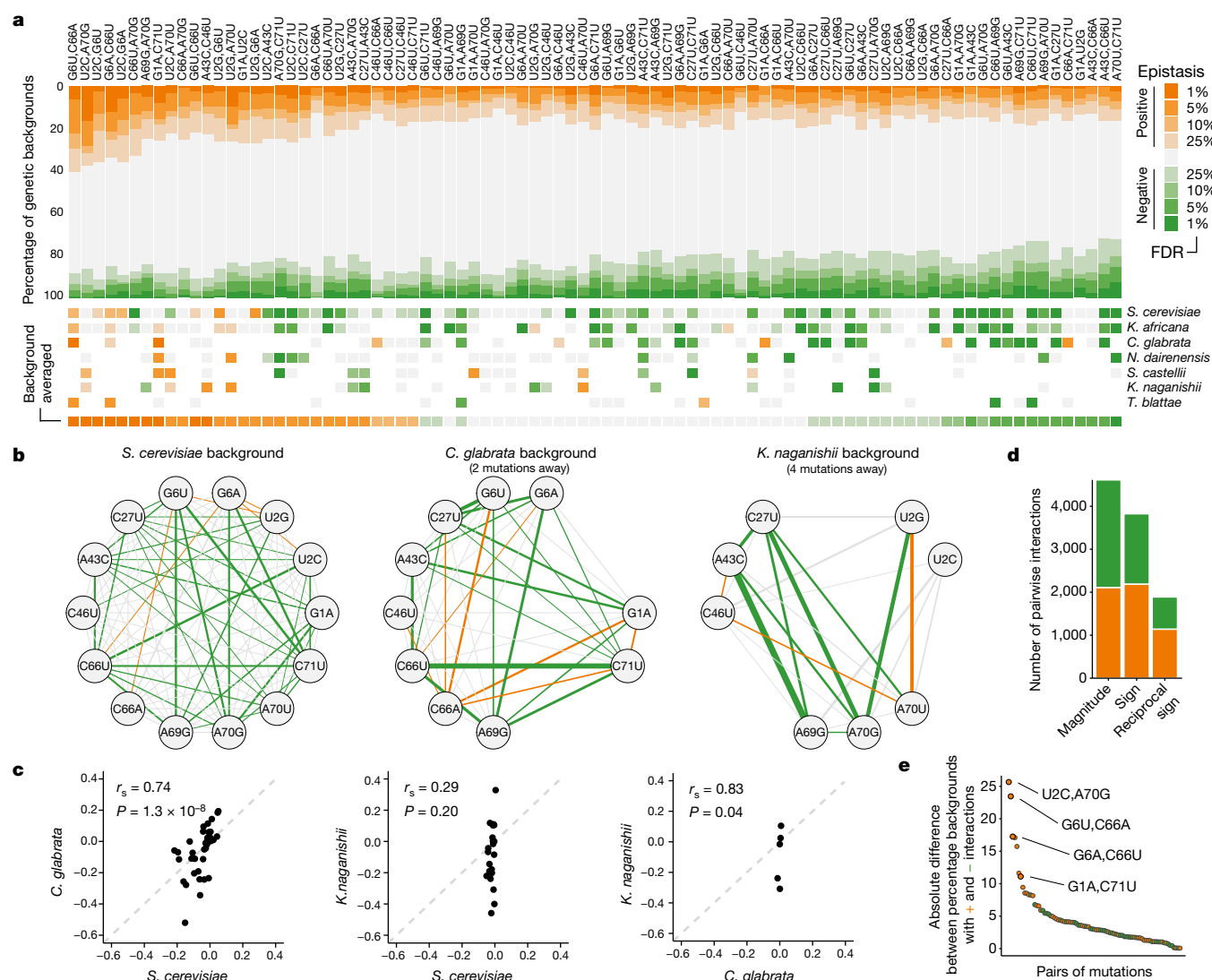
selection experiments were performed. **d**, Correlation between weighted-averaged input replicates ( $r_s$ , Spearman correlation coefficient;  $n = 4,176$  genotypes). **e**, Fitness landscape of the tRNA-Arg(CCU) genotypes (nodes). Colour indicates  $\ln(\text{fitness})$  relative to the *S. cerevisiae* tRNA. Edges connect genotypes differing by a single substitution. Genotypes and the distribution of fitness values (violins) are arranged on the  $x$ -axis according to the total number of substitutions from the *S. cerevisiae* tRNA. Highlighted nodes indicate the genotypes of the seven extant species.

Changes in base pairing only partially explained changes in the sign and magnitude of the effect of single mutations (Extended Data Fig. 6). The four pairs of mutations that restore Watson–Crick base pairs (WCBPs) were amongst the most robust positive interactions (Fig. 3e). However, even these combinations interacted negatively in a large fraction of backgrounds (5.9–8.4%). This is consistent with the presence of non-WCBP nucleotides in these positions in the tRNAs from other



**Fig. 2 | All single mutations switch sign from detrimental to beneficial in different genetic backgrounds.** **a**, The same mutation can have different fitness consequences depending on the genetic background. **b**, Significance of beneficial (blue) or detrimental (red) mutation effects in

the backgrounds of each species (left) and across all genetic backgrounds (right).  $n = 21,450$  backgrounds. **c**, Proportion of genetic backgrounds in which each mutation has beneficial (blue) or detrimental (red) effects.



**Fig. 3 | Genetic interactions between all pairs of mutations switch from positive to negative epistasis in different genetic backgrounds.** **a**, Proportion of backgrounds (top) and species (middle) in which each pair of mutations interacts positively (orange) or negatively (green) at different FDRs ( $n = 47,649$  backgrounds). Bottom, background-averaged epistasis ( $n = 87$  pairs of mutations). **b**, Interaction networks for three species (other species are shown in Extended Data Fig. 4b). Edge colours indicate epistasis sign (FDR < 0.1) and widths indicate the strength of the

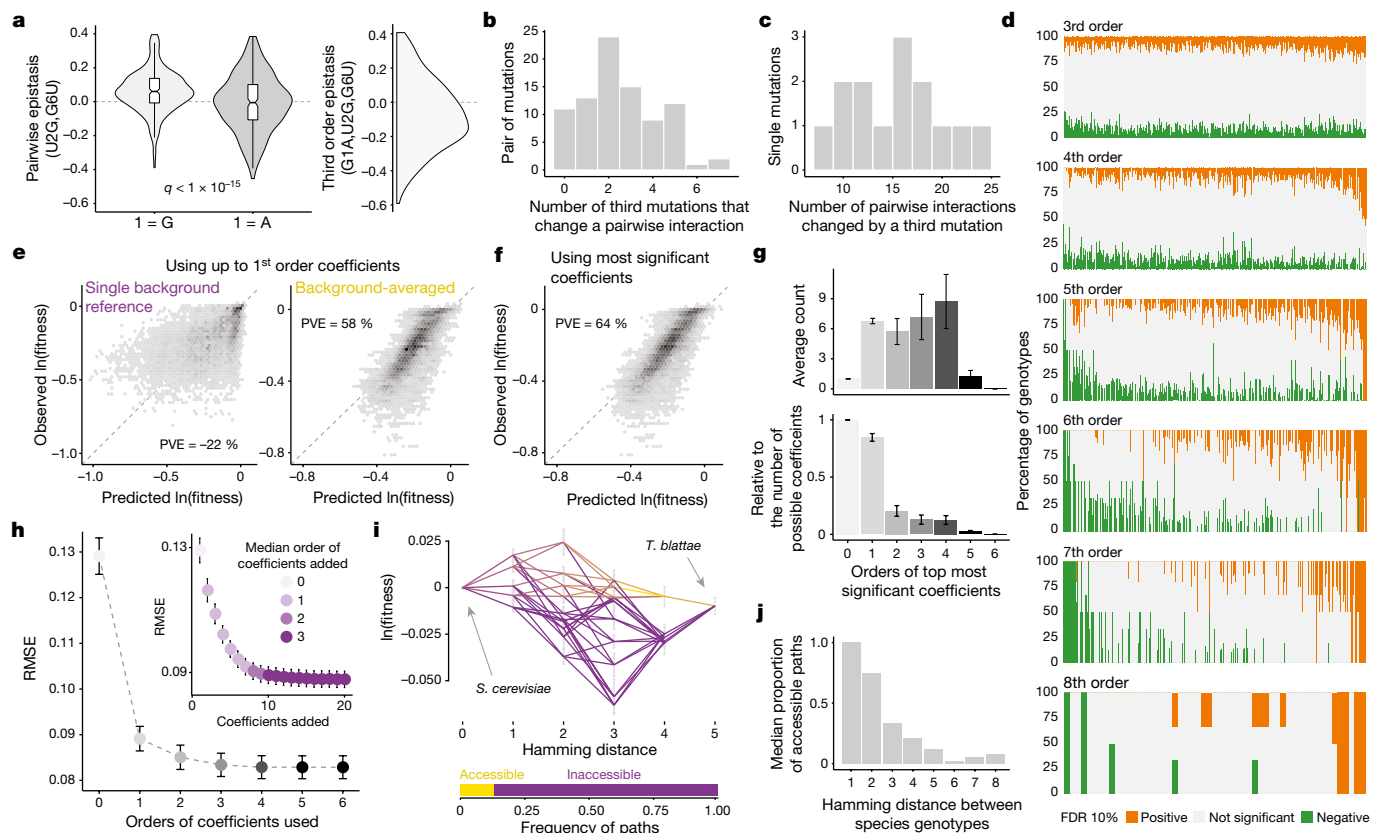
species<sup>27</sup> (Extended Data Fig. 5c). Double mutants in the same RNA strand of the acceptor stem were enriched for negative epistasis (odds ratio (OR) = 1.23, Fisher's exact test  $P = 2.15 \times 10^{-6}$ , Extended Data Fig. 5d, e) and the restoration of a WCBP was also more likely to result in a negative interaction when the stem harboured multiple additional mutations in a single strand (Extended Data Fig. 5f). This suggests that other mechanisms, for example stacking interactions, are also important determinants of tRNA function.

We next tested whether pairwise interactions changed in backgrounds containing each additional single mutation (Fig. 4a, Extended Data Fig. 7a). Notably, when averaging across genetic backgrounds, a total of 138 out of 316 possible third-order interactions were found (Extended Data Fig. 7b, FDR < 0.1), meaning that 76 out of 87 pairwise interactions were altered by the presence of a single additional mutation in the background (Fig. 4b). All 14 individual mutations altered at least eight pairwise interactions (median = 16.5, maximum = 24, Fig. 4c). As with second-order interactions, third-order interactions were enriched amongst proximal mutations and mutations found on the same strand (Extended Data Fig. 7c, d).

interaction. **c**, Comparison of epistasis scores between these three species ( $n = 43, 22$  and  $6$  comparisons from left to right, respectively). **d**, Number of positive (orange) or negative (green) magnitude, sign or reciprocal sign pairwise epistasis ( $n = 10,330$  significant interactions from 47,649 tested). **e**, Consistency of each interaction quantified as the absolute difference between the percentage of backgrounds in which the interaction is positive or negative at FDR < 0.1. Colour indicates the predominant sign. The four pairs that restore WCBPs are highlighted.

However, as for pairwise interactions, all third-order interactions (316 out of 316) also switched from positive to negative across different genetic backgrounds, indicating the presence of even higher-order epistasis (Fig. 4d). 260 out of 316 third-order interactions changed in the presence of a fourth mutation (FDR < 0.1,  $n = 740$ ). Indeed, interactions can be detected in this dataset up to the eighth order (Extended Data Fig. 7b, a total of 763 background-averaged epistatic interactions from 3,961 possible interactions tested from order one to eight, FDR < 0.1). Consistent with the behaviour of the lower-order interactions, the signs of many higher-order interactions also switch from positive to negative as the genetic background changes (Fig. 4d, 1,981 out of 3,691 interactions in the total dataset interact both positively and negatively in different genetic backgrounds at FDR < 0.1).

Finally, we evaluated the extent to which epistasis affected our ability to predict phenotypes from genotypes. We quantified the accuracy of genetic prediction in the 76 complete di-allelic sub-landscapes of eight mutations using models restricted to a single genetic background as a reference or models that averaged epistatic terms across multiple backgrounds (see Methods section 'Genetic prediction'). Although



**Fig. 4 | Averaging coefficients across genetic backgrounds and using higher order epistatic terms is important for genetic prediction.**

**a**, Changes in the distribution of pairwise epistasis when the genetic backgrounds contain or do not contain the indicated mutation (left) and the distribution of the corresponding third-order epistasis values (right). **b**, Distribution of pairwise interactions that are altered by a third mutation. **c**, Distribution of single mutations that are involved in a third order interaction. **d**, Proportion of genetic backgrounds in which each combination of mutations from third to eighth order interact positively (orange) or negatively (green) at a FDR < 0.1. **e**, Agreement between observed and predicted fitness values of all eighth order complete sub-landscapes ( $n = 19,456$  genotypes, 76 sub-landscape with 256 genotypes each) when using up to first order epistatic coefficients, relative to a single background genotype (left) or averaged across backgrounds (right, tenfold cross-validation). **f**, Agreement between observed and predicted

individual mutation effects quantified in a single genetic background provide very poor prediction (Fig. 4e), the average effect of each mutation across all genotypes within a sub-landscape improves the prediction (Fig. 4e, percentage of variance explained, PVE = 58% on held-out data, tenfold cross-validation). Including a limited number of significant interaction terms further improves the prediction (Fig. 4f, Extended Data Fig. 8a, PVE = 64%). The best models evaluated by cross-validation contain first and second order coefficients, but also higher-order interactions (Fig. 4g) that progressively reduce the prediction error (Fig. 4h). However, these models contain a relatively small number of coefficients (20 out of 256 coefficients on average across sub-landscapes, Extended Data Fig. 8b), suggesting that although pairwise and higher-order epistasis is important, reasonably sparse models can provide good genetic predictions when coefficients are measured across different genetic backgrounds.

Taken together, our results show that even single steps in sequence space substantially change the effects of both individual mutations and how these mutations combine to alter fitness. By a range of metrics, the combinatorially complete tRNA fitness sub-landscapes are most similar to rugged theoretical fitness landscapes<sup>28</sup> that constrain evolution (Extended Data Fig. 9). Indeed, the abundance of sign epistasis (Fig. 3d) limits the number of accessible evolutionary paths<sup>29</sup>, for example, paths

fitness values for all complete eighth order sub-landscapes for the best models incorporating epistatic coefficients according to the rank of their significance and evaluated by cross-validation (an average of 20 out of 256 epistatic coefficients per model). **g**, Mean orders of the most significant epistatic coefficients for the models used in f (bottom, relative to the possible number of coefficients per order; top, absolute counts). Error bars are 95% confidence intervals. **h**, Mean root-mean-square error (RMSE) across the 76 eighth order sub-landscapes when cumulatively adding the most significant coefficients determined by cross-validation (inset, colour indicates the median order of the coefficient added across the 76 sub-landscapes) or all significant coefficients from the same order (main). Error bars are 95% confidence intervals. **i**, Example of shortest paths between two extant species (top) and the accessible proportion (bottom). **j**, Average frequency of accessible paths between species.

between the genotypes of extant species (Fig. 4i, j, Extended Data Fig. 10). These results add to a growing body of evidence<sup>2</sup> that evolution is highly contingent at the molecular level. As a consequence, models that use coefficients averaged across different genetic backgrounds and that incorporate higher-order epistatic terms provide more accurate genetic prediction.

### Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0170-7>.

Received: 29 May 2017; Accepted: 9 April 2018;  
Published online: 30 May 2018

- Lehner, B. Genotype to phenotype: lessons from model organisms for human genetics. *Nat. Rev. Genet.* **14**, 168–178 (2013).
- de Visser, J. A. & Krug, J. Empirical fitness landscapes and the predictability of evolution. *Nat. Rev. Genet.* **15**, 480–490 (2014).
- Starr, T. N. & Thornton, J. W. Epistasis in protein evolution. *Protein Sci.* **25**, 1204–1218 (2016).
- Fowler, D. M. et al. High-resolution mapping of protein sequence–function relationships. *Nat. Methods* **7**, 741–746 (2010).

5. Araya, C. L. et al. A fundamental protein property, thermodynamic stability, revealed solely from large-scale measurements of protein function. *Proc. Natl Acad. Sci. USA* **109**, 16858–16863 (2012).
6. Diss, G., Lehner, B. The genetic landscape of a physical interaction. *eLife* **7**, e32472 (2018).
7. Melamed, D., Young, D. L., Gamble, C. E., Miller, C. R. & Fields, S. Deep mutational scanning of an RRM domain of the *Saccharomyces cerevisiae* poly(A)-binding protein. *RNA* **19**, 1537–1551 (2013).
8. Gong, L. I., Suchard, M. A. & Bloom, J. D. Stability-mediated epistasis constrains the evolution of an influenza protein. *eLife* **2**, e00631 (2013).
9. Olson, C. A., Wu, N. C. & Sun, R. A comprehensive biophysical description of pairwise epistasis throughout an entire protein domain. *Curr. Biol.* **24**, 2643–2651 (2014).
10. Gong, L. I. & Bloom, J. D. Epistatically interacting substitutions are enriched during adaptive protein evolution. *PLoS Genet.* **10**, e1004328 (2014).
11. Bank, C., Hietpas, R. T., Jensen, J. D. & Bolon, D. N. A systematic survey of an intragenic epistatic landscape. *Mol. Biol. Evol.* **32**, 229–238 (2015).
12. Hayden, E. J., Bendixsen, D. P. & Wagner, A. Intramolecular phenotypic capacitance in a modular RNA molecule. *Proc. Natl Acad. Sci. USA* **112**, 12444–12449 (2015).
13. Bank, C., Matuszewski, S., Hietpas, R. T. & Jensen, J. D. On the (un)predictability of a large intragenic fitness landscape. *Proc. Natl Acad. Sci. USA* **113**, 14085–14090 (2016).
14. Puchta, O. et al. Network of epistatic interactions within a yeast snoRNA. *Science* **352**, 840–844 (2016).
15. Li, C., Qian, W., Maclean, C. J. & Zhang, J. The fitness landscape of a tRNA gene. *Science* **352**, 837–840 (2016).
16. Julien, P., Miñana, B., Baeza-Centurion, P., Valcárcel, J. & Lehner, B. The complete local genotype–phenotype landscape for the alternative splicing of a human exon. *Nat. Commun.* **7**, 11558 (2016).
17. Sarkisyan, K. S. et al. Local fitness landscape of the green fluorescent protein. *Nature* **533**, 397–401 (2016).
18. Guy, M. P. et al. Identification of the determinants of tRNA function and susceptibility to rapid tRNA decay by high-throughput in vivo analysis. *Genes Dev.* **28**, 1721–1732 (2014).
19. Costanzo, M. et al. A global genetic interaction network maps a wiring diagram of cellular function. *Science* **353**, <https://doi.org/10.1126/science.aaf1420> (2016).
20. Forsberg, S. K., Bloom, J. S., Sadhu, M. J., Kruglyak, L. & Carlborg, Ö. Accounting for genetic interactions improves modeling of individual quantitative trait phenotypes in yeast. *Nat. Genet.* **49**, 497–503 (2017).
21. Tischler, J., Lehner, B. & Fraser, A. G. Evolutionary plasticity of genetic interaction networks. *Nat. Genet.* **40**, 390–391 (2008).
22. Weinreich, D. M., Lan, Y., Wylie, C. S. & Heckendorn, R. B. Should evolutionary geneticists worry about higher-order epistasis? *Curr. Opin. Genet. Dev.* **23**, 700–707 (2013).
23. Palmer, A. C. et al. Delayed commitment to evolutionary fate in antibiotic resistance fitness landscapes. *Nat. Commun.* **6**, 7385 (2015).
24. Sailer, Z. R. & Harms, M. J. Detecting high-order epistasis in nonlinear genotype–phenotype maps. *Genetics* **205**, 1079–1088 (2017).
25. Wu, N. C., Dai, L., Olson, C. A., Lloyd-Smith, J. O. & Sun, R. Adaptation in protein fitness landscapes is facilitated by indirect paths. *eLife* **5**, 16965 (2016).
26. Marcet-Houben, M. & Gabaldón, T. Beyond the whole-genome duplication: phylogenetic evidence for an ancient interspecies hybridization in the baker's yeast lineage. *PLoS Biol.* **13**, e1002220 (2015).
27. Hopf, T. A. et al. Mutation effects predicted from sequence co-variation. *Nat. Biotechnol.* **35**, 128–135 (2017).
28. Ferretti, L. et al. Measuring epistasis in fitness landscapes: The correlation of fitness effects of mutations. *J. Theor. Biol.* **396**, 132–143 (2016).
29. Weinreich, D. M., Watson, R. A. & Chao, L. Perspective: Sign epistasis and genetic constraint on evolutionary trajectories. *Evolution* **59**, 1165–1174 (2005).

**Acknowledgements** We thank J. Schmiedel for statistical guidance. This work was supported by a European Research Council Consolidator grant (616434), the Spanish Ministry of Economy and Competitiveness (BFU2011-26206 and SEV-2012-0208), the AXA Research Fund, the Bettencourt Schueller Foundation, Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR), the EMBL-CRG Systems Biology Program, and the CERCA Program/Generalitat de Catalunya. Deep sequencing was performed in the EMBL Heidelberg GeneCore Genomics Core Facility.

**Reviewer information** Nature thanks Z. Blount, D. Marks, A. Wagner and the other anonymous reviewer(s) for their contribution to the peer review of this work.

**Author contributions** J.D. performed all experiments and analyses. J.D., G.D., and B.L. designed the experiments and analyses. B.L. and J.D. wrote the manuscript.

**Competing interests** The authors declare no competing interests.

#### Additional information

**Extended data** is available for this paper at <https://doi.org/10.1038/s41586-018-0170-7>.

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41586-018-0170-7>.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>.

**Correspondence and requests for materials** should be addressed to B.L.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## METHODS

**Library design.** tRNAs orthologous to *S. cerevisiae* tRNA-Arg(CCU) (encoded by *HSX1*) were collected from the Genomic tRNA Database<sup>30</sup> or extracted from the genome of each species using BLAST<sup>31</sup> ('blastall' 2.2.25). The sequences were aligned with Clustal Omega<sup>32</sup>. Across the 12 species closest to *S. cerevisiae*, only the six species shown in Fig. 1a had substitutions in the gene, with a total of 14 substitutions in ten positions. Allowing all of these substitutions to co-occur results in a total library size of 5,184 ( $2^6 \times 3^4$ ) possible mutation combinations.

**Plasmid library construction.** An oligonucleotide of 115 nucleotides containing 72 nucleotides of tRNA flanked by 21 and 22 nucleotides of the yeast endogenous promoter and terminator was synthesized by IBA Lifesciences. At ten of the 72 positions of the tRNA, two or three different nucleotides were mixed in equal proportions during synthesis. For example, position one can be G or A, but position two can be T, G or C.

The oligonucleotide was amplified using PCR for ten cycles (Q5 Hot Start High-Fidelity DNA Polymerase, NEB), purified using an E-gel electrophoresis system (E-Gel SizeSelect Agarose Gel 2%) with column purification (MinElute PCR Purification Kit, Qiagen). Subsequently, the purified oligonucleotide was cloned into a version of the yeast centromeric plasmid pRS413 (HIS3 marker)<sup>33</sup> that contained the *HSX1* gene flanked by 218 bp of upstream and 202 bp of downstream genomic sequences (pJD001). pJD001 was linearized from the *HSX1* flanking regions (excluding the *HSX1* sequence) using PCR (Q5 Hot Start High-Fidelity DNA Polymerase, NEB) and then purified using gel extraction (QIAquick Gel Extraction Kit, Qiagen). The library of oligonucleotides was cloned into 400 µg of linearized pJD001 substituting the wild-type *HSX1* gene using a Gibson reaction (prepared in house) at 50 °C for 12 h with a ratio 5:1 of insert:vector. After dialysing the reaction with 0.025 µm VSWP membrane filters (Merck Millipore) for 1.5 h, the product was concentrated 4× using speed-vac. Six microlitres of the concentrated reaction was transformed into 100 µl of electrocompetent *Escherichia coli* (NEB 10-beta Electrocompetent *E. coli*, NEB) according to the manufacturer's protocol. Cells were allowed to recover in SOC (NEB 10-beta/Stable Outgrowth Medium) for 30 min and later transferred to 150 ml of LB medium with ampicillin 4× overnight. The total number of transformants was estimated to be  $\sim 9.59 \times 10^6$ . Given the complexity of the library, each variant was therefore represented  $\sim 1,849$  times on average. 50 ml of *E. coli* saturated culture was harvested to extract the plasmid library using plasmid midi prep (QIAfilter Plasmid Midi Kit, Qiagen).

**Selection experiment.** *Yeast strain and conditional growth defect in different environmental conditions.* The *HSX1* deletion strain was obtained by replacing the *HSX1* gene with a nourseothricin resistance cassette in the haploid laboratory strain BY4742 (MAT $\alpha$  his3 $\Delta$ 1 leu2 $\Delta$ 0 lys2 $\Delta$ 0 ura3 $\Delta$ 0 *HSX1::natMX4*) and later confirmed using colony PCR. The deletion of the single copy tRNA-Arg(CCU) (*HSX1*) in yeast was previously reported to lead to a conditional growth defect when the temperature is raised from 30 °C to 37 °C<sup>15</sup>. We found that a similar growth defect is observed if the growth medium contains high salt concentrations (1 M NaCl), and that a combination of high temperature and high salt gives an even stronger defect (Extended Data Fig. 1a). Synthetic complete medium lacking histidine (SC-HIS) 1 M NaCl at 37 °C was therefore used as the selective condition for the library selection experiment.

**Large-scale yeast transformation.** The high-efficiency yeast-transformation protocol was derived from a previously described method<sup>7</sup>. Two pre-cultures of the tRNA deletion strain were grown independently in 25 ml standard YPDA at 30 °C overnight. The next morning, the cultures were diluted into 175 ml of fresh YPDA to OD<sub>600 nm</sub> = 0.3. The two cultures were incubated at 30 °C for 4 h ( $\sim 2$ –3 generations). After the growth period, the cells were harvested and centrifuged for 5 min at 3,000g, washed in sterile water and later in SORB (100 mM LiOAc, 10 mM Tris pH 8.0, 1 mM EDTA, 1 M sorbitol). The cells were re-suspended in 8.6 ml of SORB and incubated at room temperature for 30 min. After incubation, 175 µl of 10 mg ml<sup>-1</sup> boiled salmon sperm DNA (Agilent Genomics) was added to each tube of cells, as well as 3.5 µg of plasmid library. After 10 min of gentle shaking at room temperature, 35 ml of Plate Mixture (100 mM LiOAc, 10 mM Tris-HCl pH 8, 1 mM EDTA/NaOH, pH 8, 40% PEG3350) was added to the cells and incubated at room temperature for 30 more min. 3.5 ml of DMSO was added to each tube and the cells were then heat shocked at 42 °C for 20 min (inverting tubes from time to time to ensure homogenous heat transfer). After heat shock, each independent tube of cells was centrifuged and re-suspended in 350 ml of YPD + 0.5M Sorbitol and allowed to recover for 1 h at 30 °C. The cells were then centrifuged, washed twice with SC-HIS medium and re-suspended in 350 ml SC-HIS. The two independent transformations were grown at 30 °C for  $\sim 60$  h until saturation. For the two independent transformations,  $1.5 \times 10^6$  and  $1.1 \times 10^6$  transformants were obtained, which ensured that each variant of the library was on average represented  $\sim 250$  times<sup>34</sup>.

**Competition assay.** The competition experiment had two different phases. In phase one, the environment had minimal selection on the tRNA functionality (SC-HIS at 30 °C), allowing the pool of variants to be amplified and the cells to enter the

exponential growth phase (input library)<sup>34</sup>. In the second stage, the medium was changed to a condition (SC-HIS 1 M NaCl medium at 37 °C) in which non-functional tRNA variants would lead to a severe growth defect phenotype (output library). The assay was performed immediately after yeast transformation to avoid recovering cells from frozen glycerol stocks. Once the two independently transformed cultures reached saturation ( $\sim 60$  h after plasmid transformation), they were inoculated at an OD<sub>600 nm</sub> of 0.08 in 500 ml of SC-HIS medium and grown for four generations at 30 °C ( $\sim 11$  h). When exponential phase was reached after four generations of growth, the cells were harvested and washed with selection medium (warm SC-HIS NaCl 1 M) and then inoculated in 500 ml of selection medium at an OD<sub>600 nm</sub> of 0.015. The remainder of the cells was harvested and stored at  $-20$  °C for later DNA extraction of the input libraries. Each independent input library was divided into three different output libraries (six replicates in total). Cells were grown in selective conditions for  $\sim 6.5$  generations ( $\sim 26.5$  h). This number of generations was chosen so that the average read coverage in the input would be of  $\sim 150$  reads per variant and that null alleles, which grow  $\sim 0.18$  generations every 3 h, would be detected in the output after sequencing. After 6.5 generations, the cells were harvested and the cell pellets stored at  $-20$  °C for later DNA extraction of the output libraries.

**DNA extraction and quantification.** Cell pellets (eight tubes, two inputs and six outputs) were re-suspended in 1.5 ml extraction buffer (2% Triton-X, 1% SDS, 100 mM NaCl, 10 mM Tris-HCl pH 8, 1 mM EDTA pH 8), frozen using an dry ice–ethanol bath and incubated at 62 °C in a water bath twice. Subsequently, 1.5 ml of phenol-chloro isoamyl alcohol (25:24:1 ratio, equilibrated in 10 mM Tris-HCl, 1 mM EDTA, pH8) was added, together with 1.5 g of glass beads and the samples were vortexed for 10 min. Samples were centrifuged at room temperature for 30 min at 3,200g and the aqueous phase was transferred into new tubes. The same step was repeated twice. 0.15 ml of NaOAc 3 M and 3.3 ml of cold ethanol 100% were added to the aqueous phase. The mix was incubated at  $-20$  °C for 30 min and then centrifuged for 30 min at full speed at 4 °C to precipitate the DNA. The ethanol was removed and the DNA pellet allowed to dry overnight at room temperature. DNA pellets were re-suspended in 900 µl TE 1× and treated with RNaseA (10 mg ml<sup>-1</sup>, Thermo Scientific) for 30 min at 37 °C. To desalt and concentrate the DNA solutions, a QIAEX II Gel Extraction Kit was used (75 µl of QIAEX II beads suspension). The samples were washed three times with PE buffer and eluted twice in 375 µl of 10 mM Tris-Cl buffer, pH 8.5.

**Sequencing library preparation.** The plasmid concentration in each total DNA sample was quantified in triplicate by real-time quantitative PCR, using primers that had homology to the origin of replication region of the pJD001 plasmid backbone (Supplementary Table 3). On average, we obtained  $\sim 3.5 \times 10^6$  plasmid molecules per µl of DNA sample.

A two-step PCR using high fidelity Q5 Hot Start High-Fidelity DNA Polymerase (NEB) was used to amplify the input and output libraries for sequencing. For each sample,  $\sim 150$  million plasmid molecules were amplified for ten cycles using primers with overhang homology to Illumina sequencing adapters (Supplementary Table 3). The samples were then treated with ExoSAP (Affymetrix) and cleaned using bead purification with a QIAEX II kit (10 µl pf QIAEX II beads suspension). The whole eluates, corresponding to the entire first PCR reactions, were used for the second PCR reactions (15 cycles), in which the rest of the Illumina adaptor was added as overhangs on the primers, in addition to sample-specific indexes. The DNA concentration of each individual second PCR was quantified by fluorometric quantitation (Quant-iT PicoGreen dsDNA Assay Kit) and pooled together at an equimolar ratio. Finally, the pooled sequencing library was gel purified (QIAEX II Gel Extraction Kit) and subjected to 125 bp paired-end sequencing on an Illumina HiSeq 2500v5 sequencer at the EMBL Genomics Core Facility (Heidelberg, Germany).

**From sequencing reads to fitness values.** The sequencing reads of each sample (two inputs and six outputs) were processed and filtered independently. Each sequencing read covered the entire tRNA. The 5' and 3' constant regions of the read (primers annealing sites) were removed with the 'cutadapt' software<sup>35</sup>. The forward and reverse reads were merged using PEAR<sup>36</sup> and sequences that were either not assembled owing to low quality or unexpected length were discarded. Unique genotypes were called and quantified with custom Python scripts. Genotypes with less than nine input reads in any input replicate, unexpected nucleotide substitutions (sequencing or PCR errors) or zero reads in the outputs were discarded. After filtering, we ended up with a total of 4,176 sequence genotypes quantified in all inputs and outputs.

To obtain accurate fitness and error estimates for each variant we took into account the hierarchical structure of the replicates<sup>37</sup> as well as sampling error owing to the low number of read counts<sup>38</sup>. Input and output frequencies for each genotype from each of the independent competition experiments were first calculated and then these were combined into a single output measurement for each input replicate. The number of cells expressing each genotype in each input ( $f_{in,gi}$ ) and output replicate ( $f_{out,gij}$ ) was calculated using the following formulae:

$$f_{in_{gi}} = OD_{in_{ij}} \times \frac{\text{counts}_{in_{gi}}}{\sum_{g=1}^l \text{counts}_{in_{gi}}}$$

$$f_{out_{gij}} = OD_{out_{ij}} \times \frac{\text{counts}_{out_{gij}}}{\sum_{g=1}^l \text{counts}_{out_{gij}}}$$

in which  $g$  is the genotype (from 1 to  $l$ , with  $l$  being the total number of genotypes after filtering),  $i$  is the number of input replicates (1 or 2) and  $j$  is the number of output replicates per input replicate (1 to 3).

These formulae assume that each read derives from an individual cell, so that by multiplying the frequency of reads in the output with the final ( $OD_{out}$ ) and initial culture density ( $OD_{in}$ ) we can estimate the number of cells for a particular genotype at the beginning ( $f_{in}$ ) and end ( $f_{out}$ ) of the competition experiment.

Each input and output frequency is associated to a Poisson variance given the number of read counts of each genotype and the total read count<sup>38</sup>:

$$\sigma_{in_{gi}} = \sqrt{\frac{1}{\text{counts}_{in_{gi}}} + \frac{1}{\sum_{g=1}^n \text{counts}_{in_{gi}}}}$$

$$\sigma_{out_{gij}} = \sqrt{\frac{1}{\text{counts}_{out_{gij}}} + \frac{1}{\sum_{g=1}^l \text{counts}_{out_{gij}}}}$$

We calculated a single output frequency score for each input replicate using a weighted average in which the weight of each score ( $f_{out_{gij}}$ ) is the inverse of the variance of the genotype ( $\sigma_{out_{gij}}^2$ ):

$$f_{out_{gi}} = \frac{\sum_{j=1}^3 f_{out_{gij}} \times \frac{1}{\sigma_{out_{gij}}^2}}{\sum_{j=1}^3 \frac{1}{\sigma_{out_{gij}}^2}}$$

The output frequency errors of each replicate were then combined to yield an overall output frequency error:

$$\sigma_{out_{gi}} = \sqrt{\frac{1}{\sum_{j=1}^3 \sigma_{out_{gij}}^2}}$$

The number of generations  $n_{gi}$  was then calculated as the log2 ratio of the normalized input and output frequencies:

$$n_{gi} = \log_2 \left( \frac{f_{out_{gi}}}{f_{in_{gi}}} \right)$$

with an associated error of:

$$\sigma_{n_{gi}} = \frac{1}{\ln(2)} \times \sqrt{\sigma_{out_{gi}}^2 + \sigma_{in_{gi}}^2}$$

The number of generations in each input replicate ( $n_{g1}$  and  $n_{g2}$ ) was combined using a weighted average as before to obtain a single growth measurement and an error for each genotype:

$$n_g = \frac{\sum_{i=1}^2 f_{out_{gij}} \times \frac{1}{\sigma_{n_{gi}}^2}}{\sum_{i=1}^2 \frac{1}{\sigma_{n_{gi}}^2}}$$

$$\sigma_{n_g} = \sqrt{\frac{1}{\sum_{i=1}^2 \sigma_{n_{gi}}^2}}$$

Finally, fitness values (in log-scale) relative to the *S. cerevisiae* wild type and the propagated error were calculated as follows:

$$\omega_g = \ln \left( \frac{n_g}{n_{wt}} \right)$$

$$\sigma_{\omega_g} = \sqrt{\left( \frac{\sigma_{n_g}}{n_g} \right)^2 + \left( \frac{\sigma_{n_{wt}}}{n_{wt}} \right)^2}$$

In log-space, if a particular genotype grew faster or slower than the wild type, the  $\ln(\text{fitness})$  value would be  $>0$  or  $<0$ , respectively.

**Single mutation effects, pairwise genetic interactions and higher order epistasis.** On a log-scale, the fitness effect of a mutation A on a genetic background X was calculated as the relative fitness gain of the variant AX respect to X:

$$\varepsilon_{A|X}^1 = \omega_{AX} - \omega_X$$

This fitness effect of a mutation can also be referred to as the first order epistatic term ( $\varepsilon^1$ )<sup>39</sup>.

A pairwise epistatic interaction between two mutations was defined as the difference between the observed fitness of the double mutant AB and the expected fitness obtained by the addition of the two single mutant fitness values (A and B). The fitness effects of the mutations A, B and AB can be calculated on each genetic background X by subtracting the fitness of X itself from the fitness of AX, BX and ABX, as described above. Pairwise epistasis (or second-order epistasis  $\varepsilon^2$ ) is then the change in the effect of each single mutation in the presence of the second mutation:

$$\varepsilon_{AB|X}^2 = (\omega_{ABX} - \omega_X) - ((\omega_{AX} - \omega_X) + (\omega_{BX} - \omega_X))$$

$$= \omega_{ABX} - \omega_{AX} - \omega_{BX} + \omega_X$$

$$= \varepsilon_{A|BX}^1 - \varepsilon_{A|X}^1 = \varepsilon_{B|AX}^1 - \varepsilon_{B|X}^1$$

This same analysis can be expanded to higher order terms<sup>22, 39</sup>. For example, a third-order interaction ( $\varepsilon^3$ ) is the degree to which second-order epistasis is different when a third mutation is present in the background:

$$\varepsilon_{ABC|X}^3 = \varepsilon_{AB|CX}^2 - \varepsilon_{AB|X}^2 = \varepsilon_{AC|BX}^2 - \varepsilon_{AC|X}^2 = \varepsilon_{BC|AX}^2 - \varepsilon_{BC|X}^2$$

$$= \omega_{ABCX} - \omega_{ABX} - \omega_{ACX} - \omega_{BCX} + \omega_{AX} + \omega_{BX} + \omega_{CX} - \omega_X$$

Higher order terms follow the same principle, so we can calculate any  $n^{\text{th}}$ -order term using the formula<sup>39</sup>:

$$\varepsilon^n = (-1)^0 \sum \omega^n + (-1)^1 \sum \omega^{n-1} + (-1)^2 \sum \omega^{n-2} + \dots + (-1)^n \sum \omega^{n-n}$$

$$= \sum_{i=0}^n \left( (-1)^{n-i} \sum \omega^i \right)$$

in which  $\omega^n$  are all fitness terms of order  $n$  in a specific genetic background. It is important to note that an epistatic term of any order  $n$  can only be calculated if the genotype space is complete (that is, that the fitness of all genotypes from order 0 to  $n$  were quantified in the experiment). In our dataset, higher-order epistasis was quantified up to order eight (76 cases in this dataset), which was the highest order in which the fitness of a combinatorially-complete set of genotypes could be quantified after data filtering (Extended Data Fig. 1d).

To quantify how many epistatic terms were significantly positive or negative across all the backgrounds in which they were tested, a one-sample  $t$ -test was performed (using the epistatic term and its respective propagated error). The FDR was adjusted across all the tests performed (a total of 203,240 tests for all interactions of all orders across all backgrounds) using the Benjamini–Hochberg method<sup>40</sup>.

**Controlling for background fitness, sequence divergence and the number of input sequencing reads.** Across all the data, there was a weak correlation between the fitness of the genetic background and both the fitness effect of the single mutations and pairwise epistasis (Extended Data Figs. 2c, 5a). We therefore repeated all of the analyses on the subset of the genetic backgrounds with fitness close to the wild-type *S. cerevisiae* ( $-0.15 < \text{fitness} < 0.15$ ,  $n = 1,479$  library genotypes) and also on genetic backgrounds with moderate fitness decreases ( $-0.3 < \text{fitness} < -0.15$ ,  $n = 1,577$ ). We also repeated all of the analyses on the genetic backgrounds that were closest to the *S. cerevisiae* sequence (one to four mutations away,  $n = 1,040$ ) or excluding all variants with a mean input frequency of less than 100 reads ( $n = 1,315$ ). With each of these filters we excluded approximately two thirds of the original number of variants in the library.

**Classifying pairwise epistasis.** Significant pairwise interactions in the dataset ( $n = 10,330$  out of 47,649 tested) were classified into three categories: magnitude, sign, and reciprocal sign epistasis<sup>41</sup>. Pairwise epistasis was thus classified as follows. When the fitness effect of both single mutants differs in magnitude but not in sign in the presence of the other mutation, the epistatic interaction was classified as

magnitude epistasis. For sign epistasis, the sign of one of the individual fitness has effects on changes in the presence of a second mutation. Finally, if the sign of effect changes for both individual mutations, the interaction was classified as reciprocal sign epistasis. The way a single mutation effect changes in the presence of another mutation can be inferred if the fitness effect and sign of the single mutations (A and B) and the fitness of the double mutant (AB) are known. For instance, if the two single mutations A and B have significantly beneficial (positive) effects and the double mutant has higher fitness than both single mutants, then none of the single mutations are changing sign, so this interaction would be classified as magnitude. However, if the double mutant has a fitness value lower than both single mutations, then this interaction would be classified as reciprocal sign (both single mutations are changing sign in the presence of the other). Otherwise, this interaction will be classified as sign (fitness of the double is lower than only one of the singles).

The sign of each of the single mutants in the dataset ( $n=21,450$ ) was assigned after performing a one-sample  $t$ -test (Benjamini–Hochberg FDR controlled across all tested interactions of all orders from one to eight,  $n=203,240$  as described in the Methods section ‘Single mutation effects, pairwise genetic interactions and higher order epistasis’). Single mutants with  $q \geq 0.1$  were assigned as neutral (or not-significant) and the rest as positive (beneficial) or negative (deleterious) when the fitness effect of the mutation was more or less than 0 respectively.

Exceptional interactions between two mutations in which both single mutations had a neutral category (no significant fitness effect at FDR < 0.1) were classified as magnitude epistasis (either positive or negative). When only one of the single mutations had a neutral category they were then classified as sign or magnitude epistasis depending on whether the other single mutation changed sign or not. Whenever both single mutations had either positive or negative categories, epistasis was classified as explained above.

**Background-averaged epistatic interactions.** We quantified the background-averaged epistatic interaction of a particular mutation combination (ranging from order one to eight) by averaging all epistatic coefficients of that same combination of mutations across all backgrounds in which it was found. To assess the significance of the average epistatic coefficient, the errors of all individual fitness terms were propagated and a one-sample  $t$ -test was performed. The  $P$  value was adjusted for all tests performed from order one to eight (a total of 3,691 tests) using the Benjamini–Hochberg FDR method<sup>40</sup>.

After identifying those mutations that interacted significantly when averaging across backgrounds (at FDR < 0.1), we counted the number of times the interactions between two mutations changed owing to another mutation in the background, or calculated the number of times a single mutation was able to change a pairwise interaction (Fig. 4b, c).

**Genetic prediction.** As described in the section ‘Single mutation effects, pairwise genetic interactions and higher order epistasis’, epistatic terms were calculated as linear combinations of the fitness values of genotypes of different orders. This system of linear combination can be represented in a matrix form, which allows the epistatic coefficients to be calculated from fitness values, and fitness values back from epistasis<sup>39</sup>.

In a complete  $n$  loci di-allelic genotype space, in which each locus can harbour two different nucleotides, epistatic terms can be calculated as follows:

$$\bar{\epsilon} = G\bar{\omega}$$

in which  $\bar{\omega}$  corresponds to a vector with the fitness values of the  $2^n$  genotypes from order 0 to  $n$ ,  $\bar{\epsilon}$  is a vector with all the corresponding epistatic terms and  $G$  is a matrix that defines the linear mapping between  $\bar{\omega}$  and  $\bar{\epsilon}$  for all orders.  $G$  can be recursively constructed as follows:

$$G_{n+1} = \begin{pmatrix} G_n & 0 \\ -G_n & G_n \end{pmatrix} \text{ with } G_0 = 1$$

In this case, epistatic terms are calculated relative to a single background (0<sup>th</sup> order genotype or ‘wt’). However, within a complete landscape, epistatic terms can be calculated across many different backgrounds. For instance, in a di-allelic landscape of three loci, the same single mutation effect (epistasis term of order one) can be measured four times from four different backgrounds. To obtain epistatic coefficients averaged amongst backgrounds we can use a similar version of the previous equation:

$$\bar{\epsilon} = VH\bar{\omega}$$

In this case, the  $\bar{\epsilon}$  vector corresponds to the background average epistatic coefficients.  $H$  (the Walsh–Hadamard transform<sup>22,39</sup>) defines the mapping from fitness to epistatic coefficients and can be recursively constructed as follows:

$$H_{n+1} = \begin{pmatrix} H_n & H_n \\ H_n & -H_n \end{pmatrix} \text{ with } H_0 = 1$$

The coefficient obtained by multiplying  $H$  by  $\bar{\omega}$  would correspond to the sum of the same coefficient across backgrounds, not the average. Moreover, coefficients of odd orders would have an opposite sign. The  $V$  matrix weights the coefficients by averaging and corrects the sign of odd orders depending on the order of each term.

$$V_{n+1} = \begin{pmatrix} \frac{1}{2}V_n & 0 \\ 0 & -V_n \end{pmatrix} \text{ with } V_0 = 1$$

Fitness values can be obtained by a linear combination of epistatic coefficients using the inverse mapping, for either relative or background-averaged epistatic coefficients:

$$\bar{\omega} = G^{-1}\bar{\epsilon}$$

$$\bar{\omega} = (VH)^{-1}\bar{\epsilon}$$

For an overview and extended definitions, we refer the reader to the previously published description<sup>39</sup>.

**Cross-validation.** To detect model over-fitting, we used a tenfold cross-validation approach in which the background-averaged epistatic coefficients were quantified using 90% of the genotypes (training set) within each of the 76 eight-loci tRNA sub-landscapes with the remaining 10% used for evaluation (test set). With 10% of genotypes missing, computation of seventh or eighth order coefficients is no longer possible. Coefficients of other orders were averaged across backgrounds for which all intermediate genotypes were available. To assess the significance of each epistatic coefficient, the estimates of fitness errors were propagated accordingly and the  $t$ -statistic for a one sample  $t$ -test was calculated. Within each of the ten training sets for each complete sub-landscape, the coefficients were ranked by their absolute  $t$ -statistic and cumulatively used to predict fitness of the held-out test set genotypes (least significant coefficients were iteratively set to zero before predicting fitness values) using the inverse of the Walsh–Hadamard transform as described above (using a weighting matrix  $V$  in which the weights correspond to the number of backgrounds each coefficient had been averaged across). The best predictive model for each of the ten training sets of each sub-landscape was selected as the model that gave the lowest prediction error on the corresponding test set (Extended Data Fig. 8).

The accuracy of all the above predictions was quantified using root mean square error (RMSE):

$$\text{RMSE} = \sqrt{\frac{\text{SS}_{\text{res}}}{n}}$$

in which  $\text{SS}_{\text{res}}$  is the residual sum of squares and  $n$  is the total number of predicted genotypes. To calculate the percentage of variance explained (PVE) we used the formula:

$$\text{PVE} = \left(1 - \frac{\text{SS}_{\text{res}}}{\text{SS}_{\text{total}}}\right) \times 100$$

in which  $\text{SS}_{\text{total}}$  is the total sum of squares.

**Comparisons to theoretical fitness landscapes.** We used three different landscape statistics ( $\gamma$  statistic<sup>28</sup>, roughness-to-slope ratio<sup>42</sup> and the proportion of epistasis types<sup>42</sup>) to compare the tRNA fitness landscape to theoretical landscapes. To estimate the robustness of these measurements, all the statistics were calculated for all possible di-allelic (two possible nucleotide substitutions per position) complete tRNA sub-landscapes from three to eight loci that started from the wild-type *S. cerevisiae* genotype ( $n=293, 568, 638, 403, 132, 18$  landscapes with three to eight loci respectively).

**Generation of theoretical landscapes.** We generated five different model landscapes using the software package MAGELLAN (<http://www.abi.snv.jussieu.fr/public/magellan/Magellan.main.html>): the additive model (fitness effect of each mutation is independent of the genetic background), the House of Cards model (HOC, fitness values of different genotypes are independent and identically distributed random variables), the Rough Mount Fuji model (RMF has both additive and HOC components), the Kauffman NK model (in which each locus interacts with  $K$  other loci in the landscape) and the egg box model (maximally epistatic, anti-correlated fitness landscape, in which neighbouring fitness changes systematically from low to high, or vice versa, between genetic backgrounds one step apart). Further descriptions of the models can be found in previously published works<sup>2,13,28,42</sup>. We simulated 250 di-allelic landscapes of each theoretical model of size  $n$  ( $n=3-8$ ) with an average fitness value and associated error similar to the tRNA landscape (average fitness effect of 0.04 and an associated standard error of 0.012). The RMF landscape

was modelled with a mix of 50% additive and 50% HoC and the  $K$  parameter of the NK model (each locus interacts with  $K$  loci) was set to  $K = n/2$ . These parameters were selected as they resulted in landscape statistics most similar to those of the tRNA sub-landscapes (data not shown).

**$\gamma$  statistic: correlation of fitness effects.** The  $\gamma$  statistic was recently introduced<sup>28</sup> and extended by others<sup>13</sup>.  $\gamma$  quantifies the correlation of fitness effects of the same mutation in single-mutant neighbours. It measures how the effect of a focal mutation is altered by another mutation at another locus in the background, averaged across the whole landscape. The statistic is bounded between  $-1$  and  $1$ . In a scenario without epistasis (the effect of a mutation is completely independent of the background),  $\gamma = 1$ . The  $\gamma$  measure gives information on the amount of epistasis in a combinatorially-complete landscape, but does not discriminate between different landscape topographies (two landscapes that differ in structure can have the same  $\gamma$  value). As with  $\gamma$ ,  $\gamma_d$  (the decay of correlation of fitness effects with mutational distance) can be defined as the correlation of fitness effects of mutations between genotypes that are  $1, 2, 3, \dots, d$  mutations away.  $\gamma_d$  gives extra information about the structure of the landscape, as it describes the cumulative epistatic effect of  $d$  mutations<sup>13,28</sup>. In a completely additive landscape,  $\gamma_d$  is always  $1$  because the effect of a mutation is independent of the background genotype that is  $1, 2, 3$  or  $d$  mutations away. However, in a maximally rugged fitness landscape (in which the effect of a mutation depends entirely on its genetic background)  $\gamma_1$  is  $0$  and  $\gamma_d$  is  $0$  for all values of  $d$ . The behaviour of  $\gamma_d$  as a function of  $d$  varies for different theoretical landscape models<sup>13,28</sup> (Extended Data Fig. 9a).

We calculated  $\gamma_d$  values for all possible complete di-allelic tRNA sub-landscapes of three to eight mutations combinations that contained the *S. cerevisiae* genotype using the software MAGELLAN (eight being the maximum number of loci in which a complete genotype space is available in the dataset). We later compared the statistic to the values for the theoretical landscapes. As a measure of similarity, we calculated the Euclidean distance between the  $\gamma_d$  of all tRNA sub-landscapes and the  $\gamma_d$  of the theoretical models (each tRNA landscape,  $n = 73,250, 142,000, 159,500, 100,750, 33,000$  and  $4,500$  for tRNA landscapes from three to eight mutations respectively, was compared to the 250 simulations of each theoretical landscape).

**Other quantitative measures of landscape ruggedness.** In addition to the  $\gamma$  statistic, for all complete tRNA and theoretical sub-landscapes from three to eight loci, we also calculated the roughness-to-slope ratio ( $r/s$  ratio) and characterized the local pairwise epistatic interactions. The  $r/s$  ratio measures how well the landscape can be described by a linear model, which corresponds to the purely additive limit<sup>42</sup>. The roughness is given by the variance of the residuals from the linear model and the slope is given by the average of the absolute values of the linear coefficients. The higher the  $r/s$ , the higher the deviation from the linear model and the more epistasis is present (in a non-epistatic scenario,  $r/s = 0$ ). To characterize the local interactions of each landscape we calculated the fraction of magnitude, sign or reciprocal sign pairwise epistasis within each landscape. We used the software MAGELLAN to calculate all the described statistics.

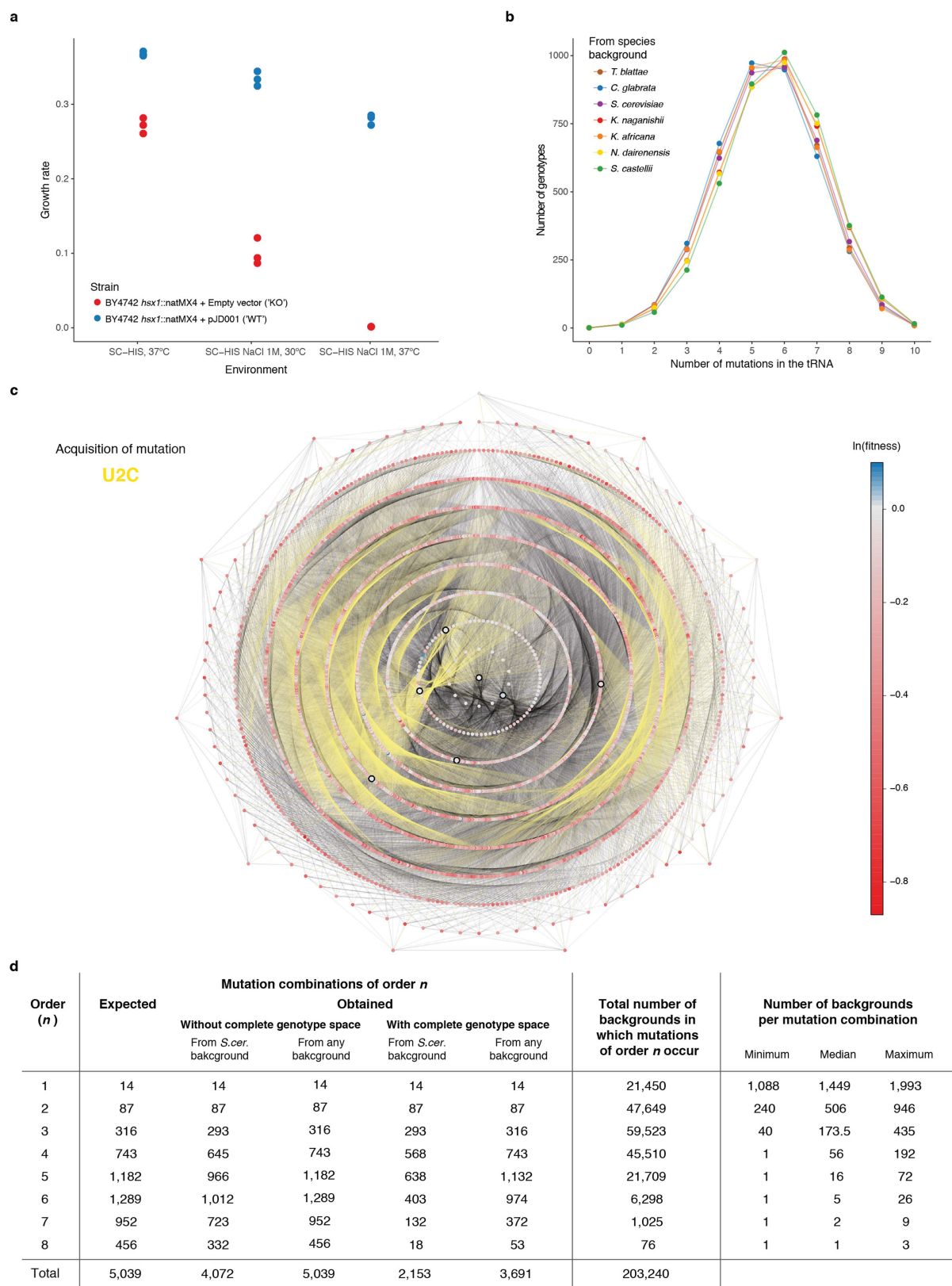
**Accessible paths between extant species.** An accessible path between two genotypes in the landscape was defined as a mutation trajectory in which none of the

intermediate genotypes has significantly lower fitness than both the initial and final genotypes that they connect ( $t$ -test between all the intermediate genotypes against the origin and end-point genotypes,  $n = 1-8$  tests). A path that had at least one deleterious intermediate genotype ( $P < 0.05$ ) was classified as inaccessible. We measured the number of accessible direct (shortest) paths between 20 pairwise comparisons of the extant genotypes in the landscape using the R package igraph. **Statistical analyses.** All statistical analyses were performed in R (v.3.3.3) and figures were made using the R package ggplot2. Lower and upper hinges of box plots correspond to the first and third quartiles (25<sup>th</sup> and 75<sup>th</sup> percentiles). The upper and lower whiskers extend from the hinge to the largest and lower value no further than  $1.5 \times \text{IQR}$  (inter-quartile range) respectively. Higher or lower points (outliers) are plotted individually (or not plotted in those cases where the box plot is plotted together with a violin plot). Notches give roughly 95% confidence interval for comparing the medians.

**Reporting summary.** Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

**Data availability.** The complete dataset is available as Supplementary Table 1. Custom code used in this study is available from the authors upon request. Raw sequencing data has been submitted to GEO (accession number GSE99418).

30. Chan, P. P. & Lowe, T. M. GtRNAdb: a database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Res.* **37**, D93–D97 (2009).
31. Altshul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
32. McWilliam, H. et al. Analysis tool web services from the EMBL-EBI. *Nucleic Acids Res.* **41**, W597–W600 (2013).
33. Sikorski, R. S. & Hieter, P. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. *Genetics* **122**, 19–27 (1989).
34. Matuszewski, S., Hildebrandt, M. E., Ghenu, A. H., Jensen, J. D. & Bank, C. A statistical guide to the design of deep mutational scanning experiments. *Genetics* **204**, 77–87 (2016).
35. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal* **17**, 10–12, (2011).
36. Zhang, J., Kobert, K., Flouri, T. & Stamatakis, A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* **30**, 614–620 (2014).
37. Crawley, M. J. *The R Book*. (Wiley, Chichester, 2007).
38. Rubin, A. F. et al. A statistical framework for analyzing deep mutational scanning data. *Genome Biol.* **18**, 150 (2017).
39. Poelwijk, F. J., Krishna, V. & Ranganathan, R. The context-dependence of mutations: a linkage of formalisms. *PLOS Comput. Biol.* **12**, e1004771 (2016).
40. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B Stat. Methodol.* **57**, 289–300 (1995).
41. Poelwijk, F. J., Kiviet, D. J., Weinreich, D. M. & Tans, S. J. Empirical fitness landscapes reveal accessible evolutionary paths. *Nature* **445**, 383–386 (2007).
42. Szendro, I. G., Schenk, M. F., Franke, J., Krug, J. & de Visser, J. A. Quantitative analyses of empirical fitness landscapes. *J. Stat. Mech.* **2013**, P01005 (2013).



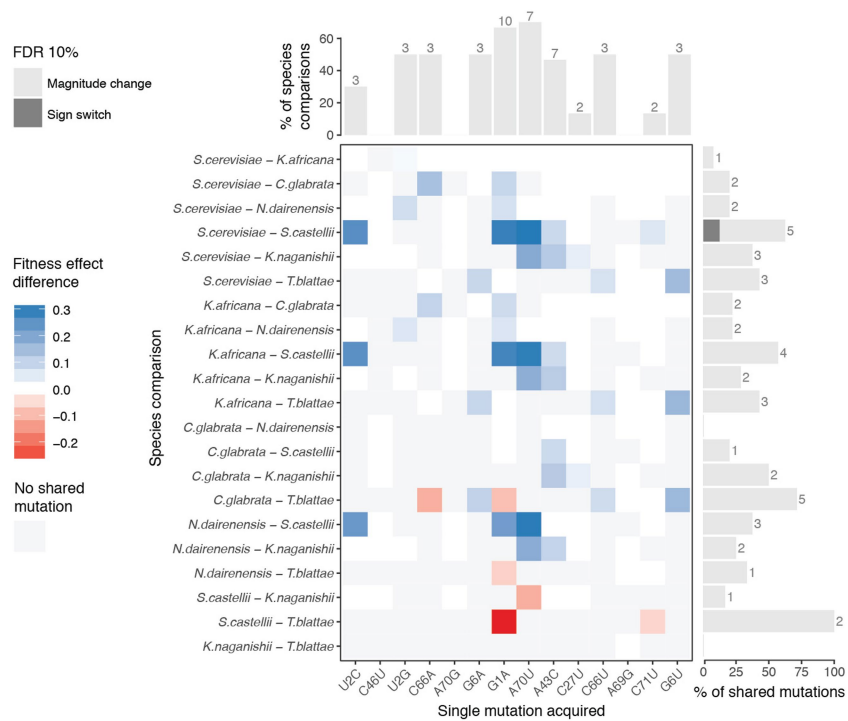
Extended Data Fig. 1 | See next page for caption.

**Extended Data Fig. 1 | Experimental design.** **a**, Maximum growth rate (measured in a plate reader using spectrophotometry) of tRNA-Arg(CCU) (*HSX1*) deletion strain carrying either an empty plasmid (red) or a single-copy plasmid expressing wild-type tRNA-Arg(CCU) (blue) at high temperature, high salt, and high temperature with high salt ( $n = 3$  independent colonies from the plasmid transformation).

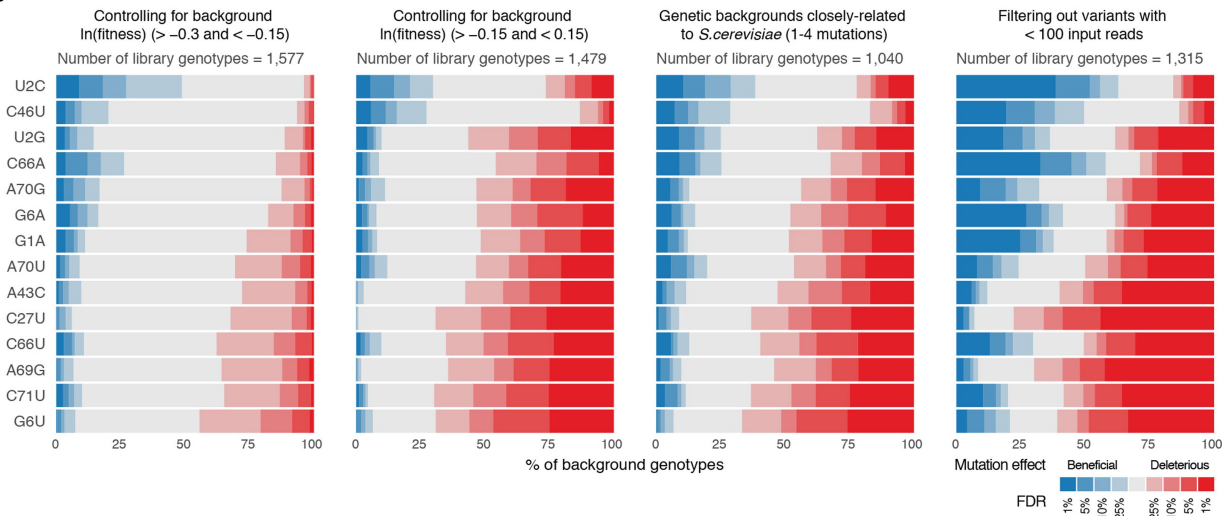
**b**, Distribution of number of mutations per genotype in the library relative to the sequence of the tRNA from each species. **c**, Genotype network of the 4,176 tRNA-Arg(CCU) variants. Each node is one genotype. Colour indicates the  $\ln(\text{fitness})$  relative to *S. cerevisiae*. Edges connect genotypes differing by a single substitution, acquisition of a U2C mutation is highlighted in yellow as example. Genotypes are arranged in concentric

circles according to the total number of substitutions (one to ten) from the *S. cerevisiae* tRNA, which is the central node. Highlighted nodes indicate the genotypes of the seven extant species. **d**, Table showing the possible number of mutation combinations from order one to eight, with or without a complete genotype space (whether all intermediate genotypes are measured in the library or not) when using *S. cerevisiae* as a reference or any other background (the effect of a given combination of mutations can be measured from at least one genetic background). The total number of unique backgrounds is also indicated, together with the minimum, median and maximum number of backgrounds in which these mutations can be found.

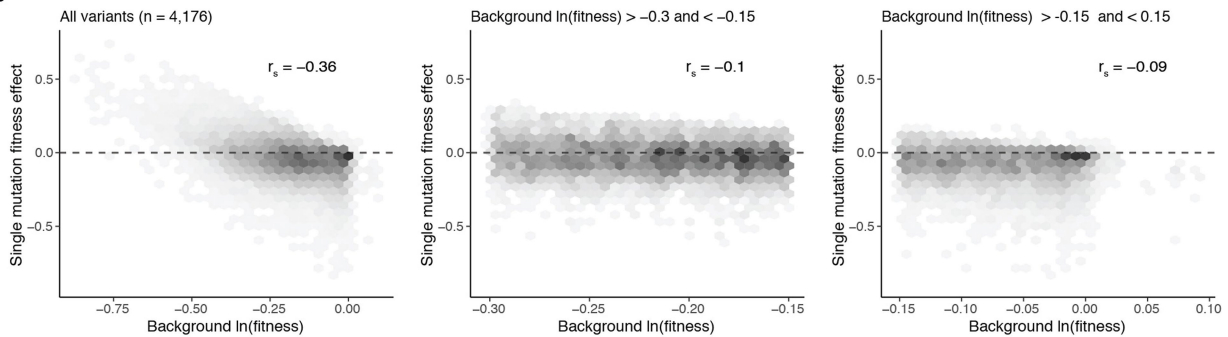
**a**



**b**



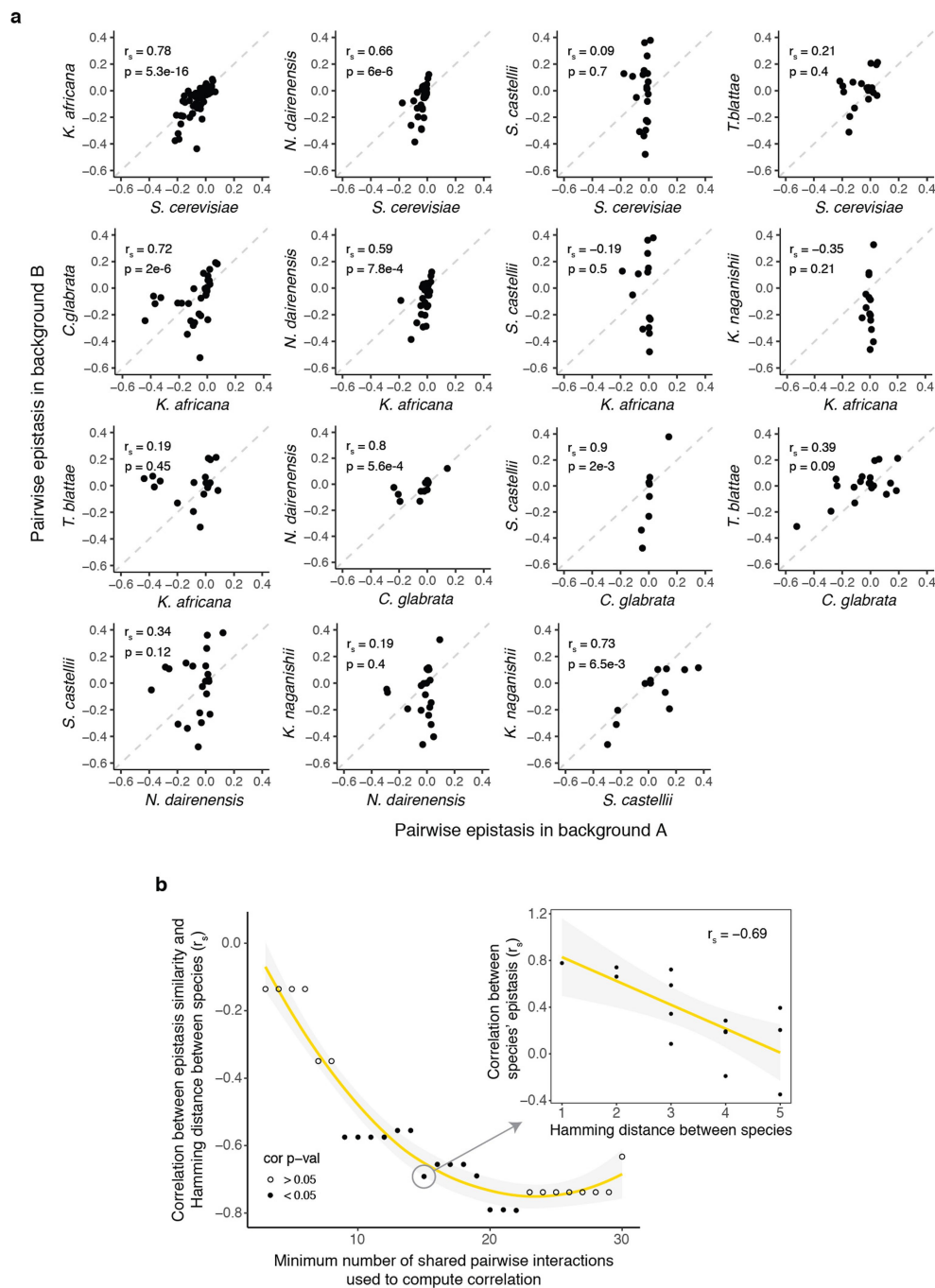
**c**



Extended Data Fig. 2 | See next page for caption.

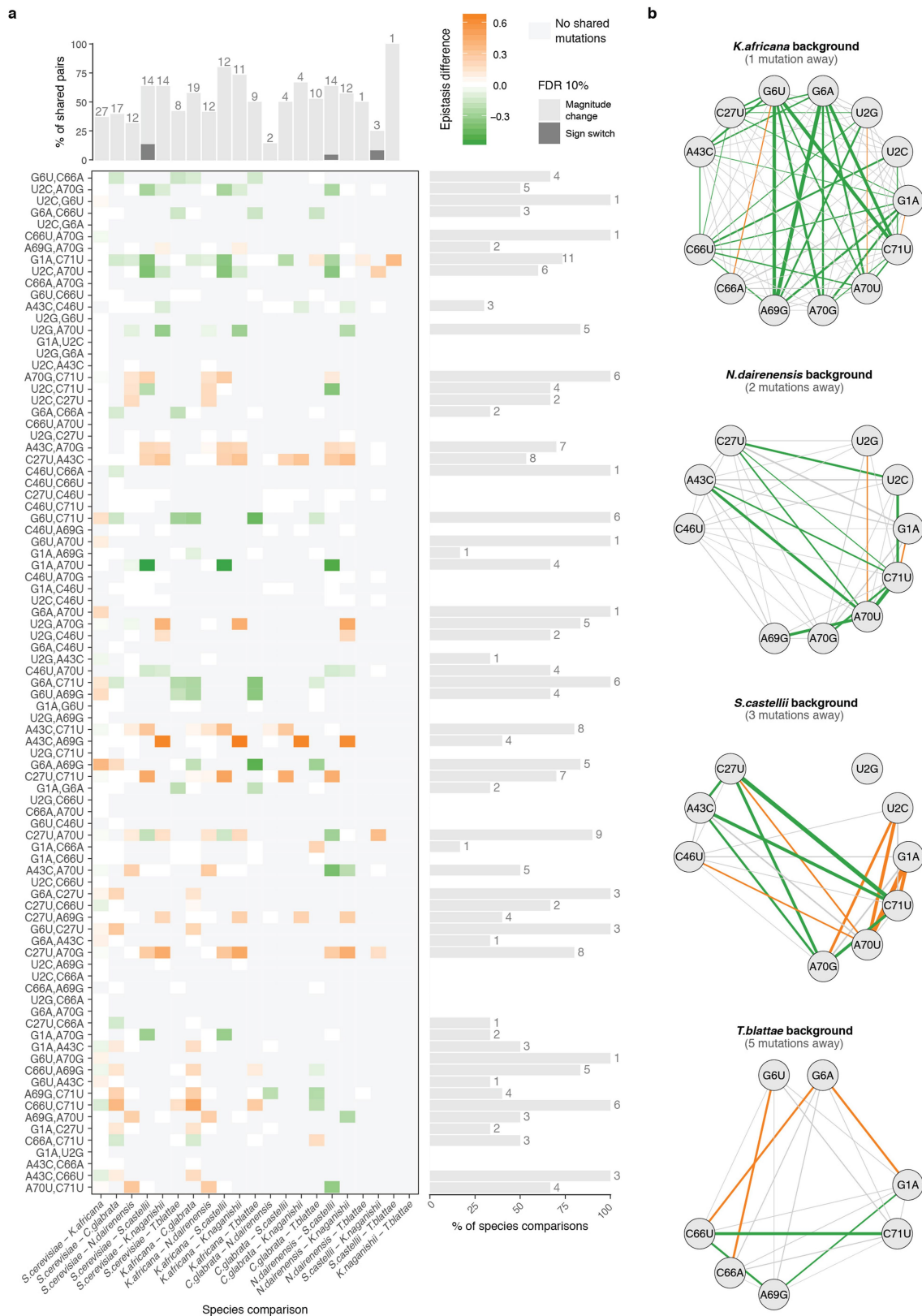
**Extended Data Fig. 2 | Mutations have varying fitness effects in different backgrounds.** **a**, Single mutations (columns) have effects that differ significantly between genetic backgrounds from different species (rows). Paired two-sided *t*-test between fitness effects of mutations of tRNAs from different species (145 tests of  $n = 6$ ). Significant fitness effects differences ( $\text{FDR} < 0.1$ ) shown in blue (positive) or red (negative), non-significant differences ( $\text{FDR} \geq 0.1$ ) coloured in white. Mutations that were not shared are coloured in grey (that is, a substitution that would result in a mutation in one species but is part of the wild-type background in another). Bar plots show the percentage (absolute numbers on top) of species comparisons or shared mutations between species in which the effect of the mutation significantly changes in magnitude (light grey) or switches sign (dark grey). **b**, Proportion of genetic backgrounds in which

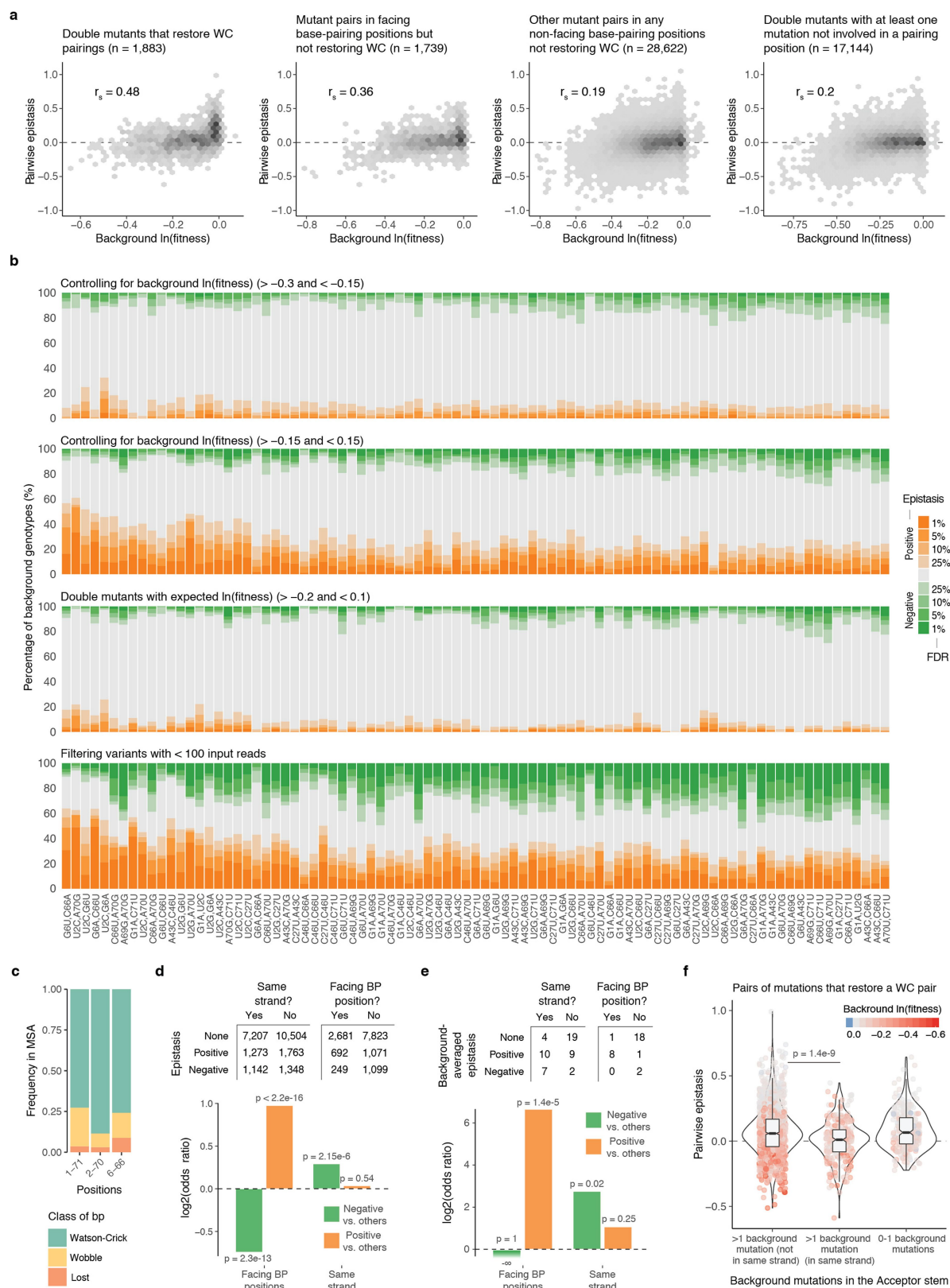
each mutation has a beneficial (blue) or detrimental (red) fitness effect at different FDRs for backgrounds with  $-0.3 < \ln(\text{fitness}) < -0.15$  (left), backgrounds with  $-0.15 < \ln(\text{fitness}) < 0.15$  (middle left), genotypes with no more than four mutations from the *S. cerevisiae* sequence (middle right) and genotypes with average input read counts of more than 100 (right). *q* values were obtained after adjusting for FDR across the total number of single mutations with unique background after filtering ( $n = 10,746, 6,129, 3,568, 6,338$  tests respectively). **c**, Fitness effect of single mutations plotted against the  $\ln(\text{fitness})$  of the backgrounds in which the mutation are made; for all genetic backgrounds (left), backgrounds with  $-0.3 < \ln(\text{fitness}) < -0.15$  (middle) and backgrounds with  $-0.15 < \ln(\text{fitness}) < 0.15$  (right).



**Extended Data Fig. 3 | Comparison of epistasis scores between all pairs of species. a**, Comparison of epistasis scores for species pairs not shown in Fig. 3c. Pairs of species that share less than three mutations are not shown. **b**, Decline of correlation between epistasis scores and Hamming distance

between the tRNA genotypes from different species (inset). The left plot shows how this negative correlation holds when restricting the minimum number of shared pairs of mutations between the two species to compute the correlation.



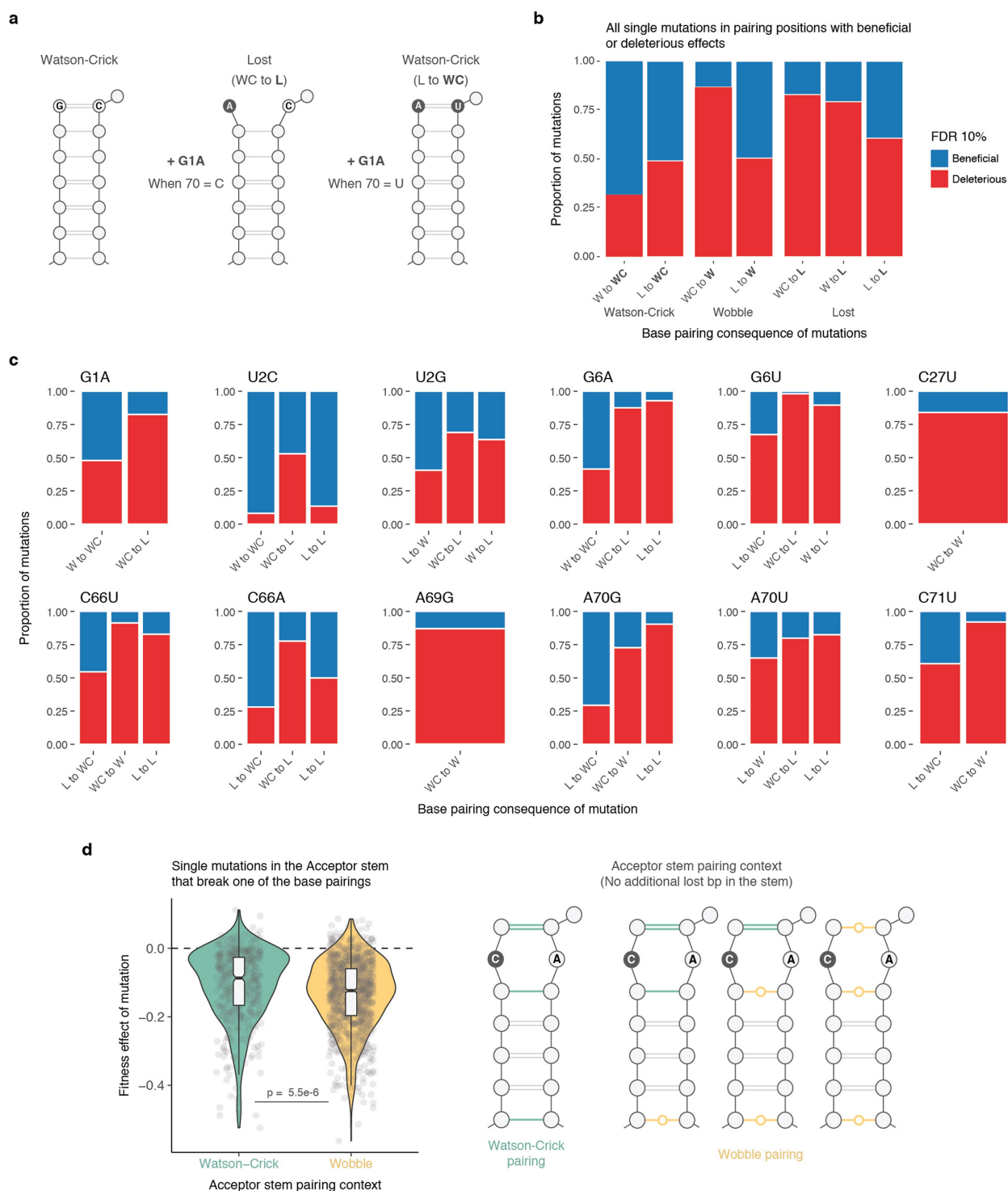


Extended Data Fig. 5 | See next page for caption.

**Extended Data Fig. 5 | Pairwise epistatic interactions switch from positive to negative.**

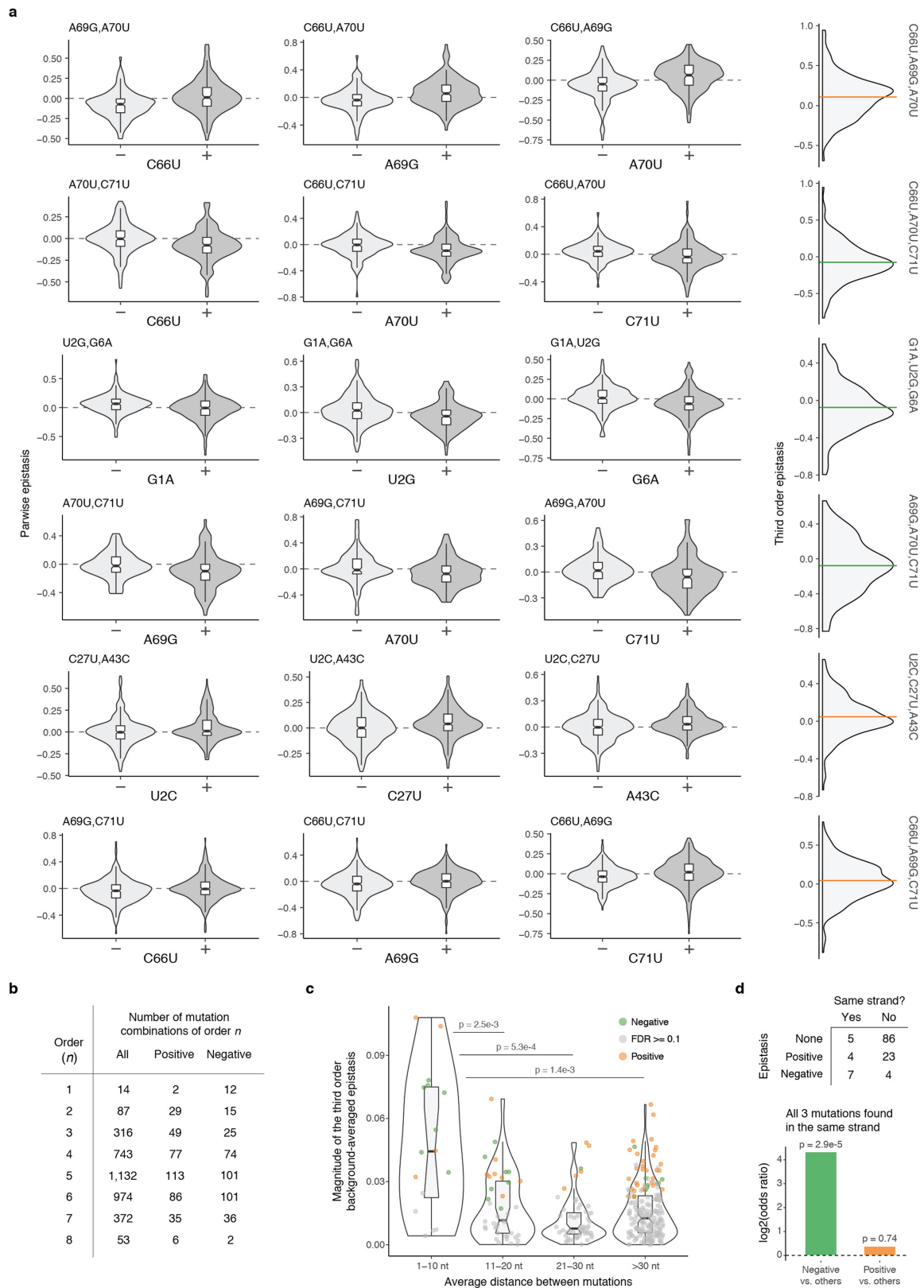
**a**, Epistasis scores between pairs of mutations plotted against the  $\ln(\text{fitness})$  of the genetic background. Scatter plots are divided into double mutants that restore WCBPs (left,  $n = 1,883$ ), other double mutants in which both mutation are in facing base pair positions (middle left,  $n = 1,739$ ), in base pair positions but not facing each other (middle right,  $n = 28,622$ ), and the rest (right,  $n = 17,144$ ). **b**, Proportion of genetic backgrounds in which each pair of mutations interacts with positive (orange) or negative (green) epistasis at different FDRs restricted to genetic backgrounds with  $-0.3 < \text{fitness} < -0.15$  (top), with  $-0.15 < \text{fitness} < 0.15$  (top middle), with additive expected fitness outcome greater than  $-0.2$  and less than  $0.1$  (middle bottom) or when excluding all genotypes with average input counts less than 100 (bottom). 23,128, 23,652, 29,628 and 15,306 one sample two-sided  $t$ -tests ( $n = 6$ ). **c**, A small fraction of tRNA-Arg(CCU) from other eukaryotic species have lost the base pairing in positions 1–71, 2–70 and 6–66 of the tRNA (multiple sequence alignment (MSA) across 1,614 species was taken from previously published work<sup>27</sup>; sequences with indels were excluded).

**d**, Number of positive, negative or not significant pairwise interactions at  $\text{FDR} < 0.1$  within the acceptor stem of the tRNA ( $n = 23,237$ ) when both mutations are found in the same helix strand or when each mutation is located in a different strand ( $n = 13,615$ ).  $\log_2$  odds ratio shown below together with two-sided Fisher's exact test  $P$  values. **e**, Number of positive, negative and non-significant background-averaged pairwise interactions between pairs of mutations in the acceptor stem that are found in the same RNA strand and between mutations that are in positions that base pair with each other.  $\log_2$  odds ratio and two-sided Fisher's exact test  $P$  values are shown below. **f**, Distribution of pairwise epistasis values of mutation pairs that restore a canonical WCBP depending on the location of their background mutations in the acceptor stem ( $P$  values from Welch's two-sided  $t$ -test,  $n = 263$  or  $n = 1,368$  when more than one background mutations are in the same strand or not, respectively). The same result is obtained when epistasis values are corrected for the  $\ln(\text{fitness})$  of the background (residuals of a linear model using background  $\ln(\text{fitness})$  to predict epistasis, data not shown).



**Extended Data Fig. 6 | Changes in base pairing partially explain the consequences on fitness of single mutations.** **a**, A single mutation can either disrupt or restore a canonical WCBP depending on the background context. **b**, Percentage of deleterious or beneficial single mutations (at FDR < 0.1) that restore or disturb a canonical WCBP in any base pairing position of the tRNA. From a total of 4,300 mutations that restore WCBP, 721 are beneficial and 498 deleterious. 13,195 mutations result in the loss of a canonical pair ( $n = 6,806$  mutations that create a wobble base pair and  $n = 6,389$  that completely break the base pair interaction), of these 3,030 and 721 have significant deleterious and beneficial effects,

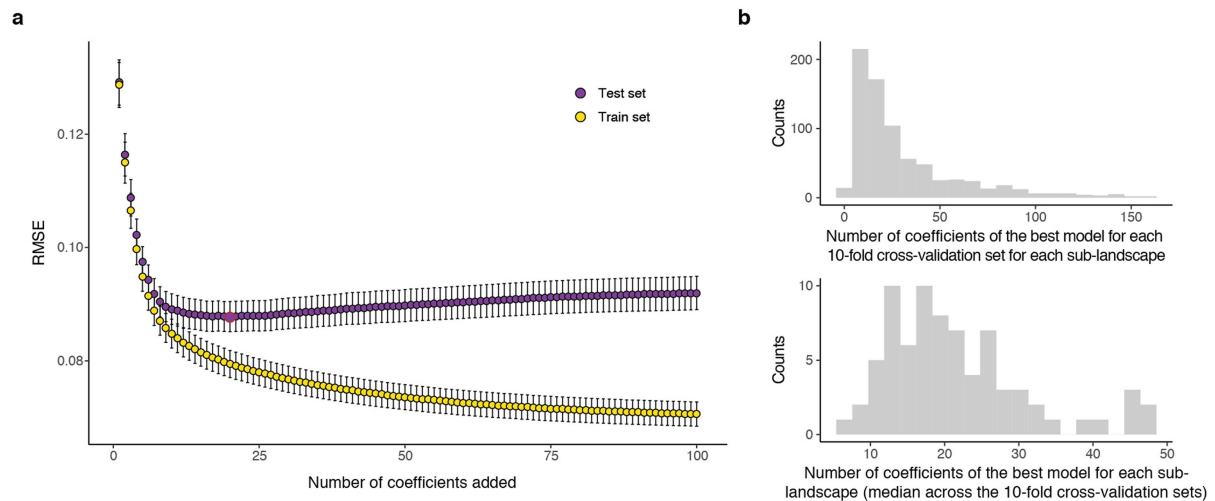
respectively. WC, Watson-Crick, W, wobble and L, lost base pair. **c**, Same as **b** but split by mutation identity. **d**, Distribution of the effects of mutations in the tRNA acceptor stem that break a base pairing (left,  $n = 1,356$  single mutations with higher background fitness than  $-0.15$ ) have more deleterious effects when the neighbour base-pairing positions are composed of one or more wobble interactions ( $n = 921$ ), instead of all canonical WCBP ( $n = 435$ , average fitness effect difference = 0.028, Welch's two-sided  $t$ -test  $P$  value shown). Right plot illustrates the context of the base pairing of the stem.



Extended Data Fig. 7 | See next page for caption.

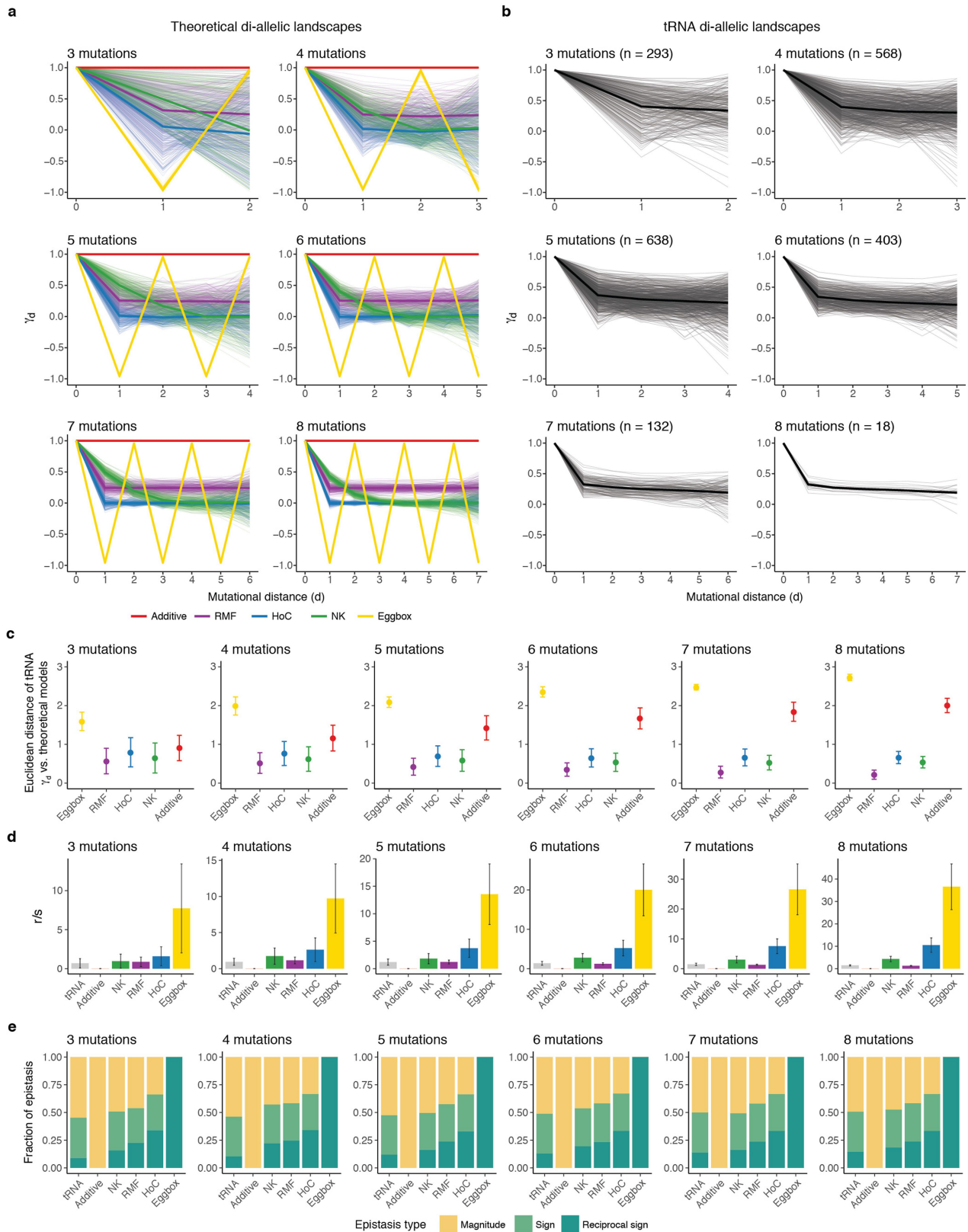
**Extended Data Fig. 7 | Background-averaged third and higher-order interactions.** **a**, The most significant background-averaged third-order interactions (8 out of 74,  $FDR < 0.1$ ,  $n = 3,691$  tests for all interactions across all orders). The first three plots of each row show how the distribution of pairwise epistasis of two mutations across different genetic backgrounds (each double mutation can be found in a median of 506 different genetic backgrounds) changes in the presence or absence of a third mutation. The paired differences between pairwise interactions in those three cases correspond to third order epistatic coefficients. Distributions of third-order epistasis for the same three mutations are shown to the right. Horizontal lines correspond to the background-averaged third-order epistatic term, coloured by sign (orange or green for positive or negative respectively). **b**, Number of significantly positive and negative background-averaged epistatic interactions of order one

to eight (at  $FDR < 0.1$ ). **c**, Distribution of the absolute magnitude of averaged third-order interactions plotted against the mean nucleotide distance between the three mutations ( $n = 316$  triple mutations). Welch's two-sided  $t$ -test  $P$  values for differences between the groups are shown. Significant interactions (one-sample two-sided  $t$ -test at  $FDR < 0.1$ ) are coloured in orange or green for positive or negative epistasis respectively. **d**, Top, Number of positive, negative or non-significant background-averaged third-order interactions ( $FDR < 0.1$ ) within the acceptor stem of the tRNA when both mutations are found in the same helix strand or not ( $n = 129$ ). Bottom, the log2 odds ratios (when all three mutations are found in the same strand of the tRNA acceptor stem) of significantly positive interactions versus others (negative or not significant interactions) and significantly negative interactions versus other double mutants.  $P$  values reported from the two-sided Fisher's exact test.



**Extended Data Fig. 8 | Genetic prediction. a,** Mean RMSE of the fitness prediction for tenfold cross-validation held-out genotypes (purple, test set) or genotypes included in the training set (yellow) for each of the eight-mutation sub-landscapes when progressively adding the 100 most significant epistatic coefficients out of the 256 possible coefficients. Highlighted in red is the average number of epistatic coefficients to

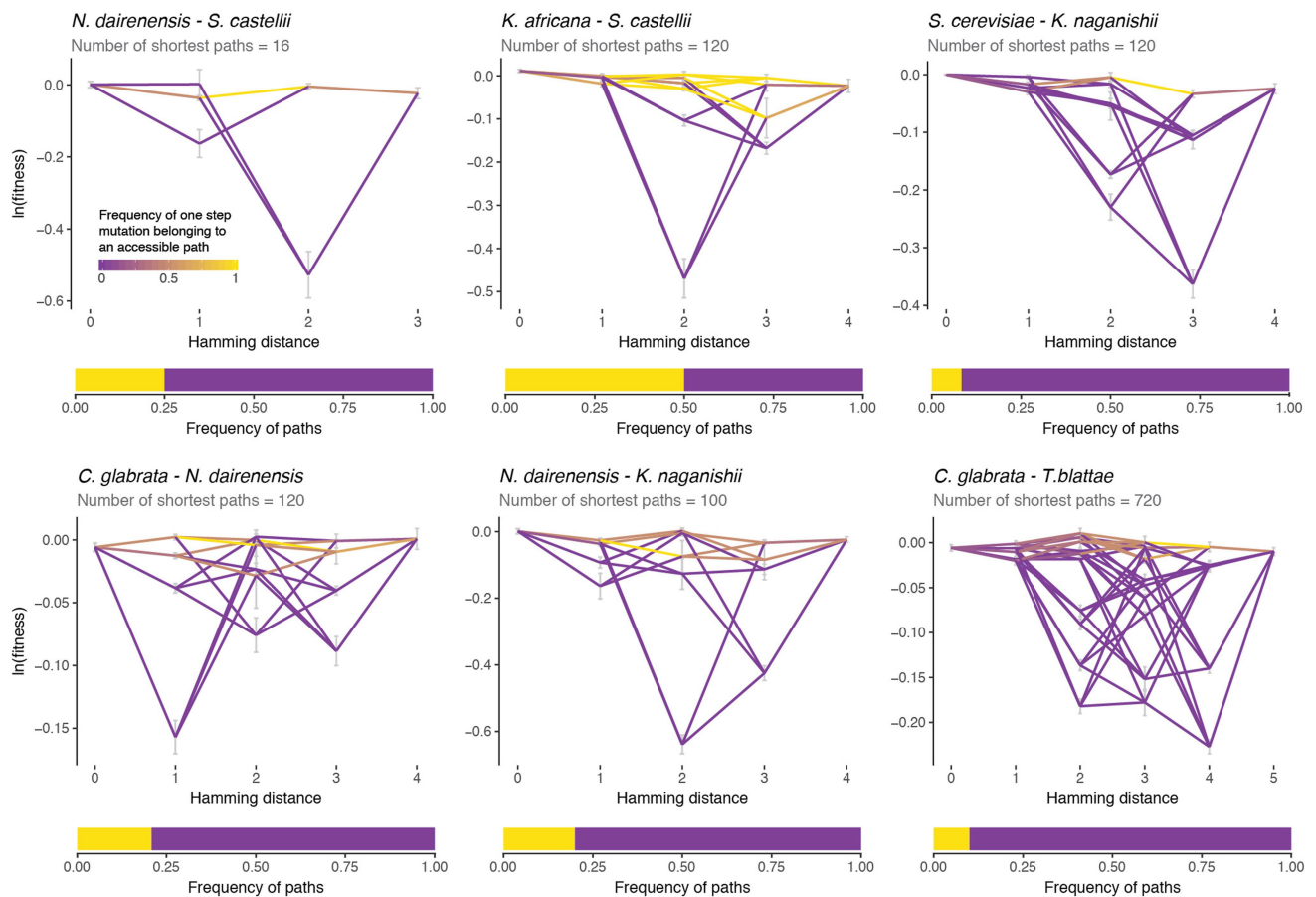
obtain the lowest RMSE across all the sub-landscapes. **b,** Histogram of the minimum number of epistatic coefficients that give the minimum RMSE when predicting the fitness of the test genotypes by tenfold cross-validation in all complete eight-mutation sub-landscapes (top). Histogram of the median number of coefficients for each sub-landscape (bottom).



Extended Data Fig. 9 | See next page for caption.

**Extended Data Fig. 9 | Comparison of the combinatorially-complete tRNA sub-landscapes to theoretical fitness landscapes.** **a**, Expected pattern of the average correlation of fitness effects  $\gamma_d$  at different mutational distances for theoretical di-allelic fitness landscapes with three to eight mutated positions. The average  $\gamma_d$  behaviour is highlighted in bold for each theoretical landscape ( $n = 250$  simulated landscapes for each theoretical model). The NK landscape was modelled with  $K = L/2$  ( $L$ , number of mutated positions) and the RMF as a mixture of 50% additive and 50% HoC. **b**, Decay of  $\gamma_d$  with mutational distance for all tRNA complete di-allelic sub-landscapes containing the *S. cerevisiae* parental genotype of three to eight loci (mean behaviour of  $\gamma_d$  in bold).

**c**, Mean euclidean distance between the  $\gamma_d$  for the tRNA sub-landscapes and the  $\gamma_d$  of theoretical landscapes (each tRNA landscape was compared to the 250 simulations of each theoretical landscape,  $n = 73,250, 142,000, 159,500, 100,750, 33,000$  and  $4,500$  for tRNA landscapes from three to eight mutations respectively). **d**, **e**, Mean roughness-to-slope ratio ( $r/s$ ) (**d**) and epistasis classes (**e**) for all combinatorially-complete tRNA di-allelic landscapes from three to eight mutations, as well as for all theoretical landscape models ( $n = 250$  for each theoretical landscape models and 293, 568, 638, 403, 132 and 18 tRNA landscapes from three to eight mutations respectively). Error bars are s.d.



**Extended Data Fig. 10 | Direct paths accessibility between extant species.** Shortest paths between some pairs of extant species (top) together with the proportion of them that are accessible (bottom; yellow, accessible; purple, inaccessible). Nodes are the  $\ln(\text{fitness})$  of the species genotypes and the intermediate genotypes between them. Edge colours indicate the

frequency at which a one-step mutation belongs to an accessible path (completely accessible, yellow; completely inaccessible, purple). Error bars are  $\ln(\text{fitness})$  s.e.m. of each genotype (propagated error from the  $n=6$  replicates).

# Molecular tuning of electroreception in sharks and skates

Nicholas W. Bellono<sup>1,2,3\*</sup>, Duncan B. Leitch<sup>1,3</sup> & David Julius<sup>1\*</sup>

**Ancient cartilaginous vertebrates, such as sharks, skates and rays, possess specialized electrosensory organs that detect weak electric fields and relay this information to the central nervous system<sup>1–4</sup>. Sharks exploit this sensory modality for predation, whereas skates may also use it to detect signals from conspecifics<sup>5</sup>. Here we analyse shark and skate electrosensory cells to determine whether discrete physiological properties could contribute to behaviourally relevant sensory tuning. We show that sharks and skates use a similar low threshold voltage-gated calcium channel to initiate cellular activity but use distinct potassium channels to modulate this activity. Electrosensory cells from sharks express specially adapted voltage-gated potassium channels that support large, repetitive membrane voltage spikes capable of driving near-maximal vesicular release from elaborate ribbon synapses. By contrast, skates use a calcium-activated potassium channel to produce small, tunable membrane voltage oscillations that elicit stimulus-dependent vesicular release. We propose that these sensory adaptations support amplified indiscriminate signal detection in sharks compared with selective frequency detection in skates, potentially reflecting the electroreceptive requirements of these elasmobranch species. Our findings demonstrate how sensory systems adapt to suit the lifestyle or environmental niche of an animal through discrete molecular and biophysical modifications.**

Electrosensory cells from the little skate (*Leucoraja erinacea*) express specialized low threshold  $\text{Ca}_v1.3$  voltage-gated calcium ( $\text{Ca}^{2+}$ ) channels and big-conductance potassium ( $\text{K}^+$ , BK) channels that functionally couple to produce cellular membrane voltage oscillations<sup>6,7</sup>. In electrosensory cells from the chain catshark (*Scyliorhinus retifer*, Fig. 1a), we similarly observed voltage-activated inward calcium currents ( $I_{\text{CaV}}$ ) that were sensitive to L-type voltage-gated  $\text{Ca}^{2+}$ -channel ( $\text{Ca}_v$ ) modulators, had a low voltage threshold for activation, had steep voltage-dependence, and had a slow inactivation profile that contributed to a large window-current across physiological membrane voltages (Extended Data Fig. 1a–d, f). As with skates, the pore-forming  $\alpha$  subunit of  $\text{Ca}_v1.3$  was the predominant  $\text{Ca}_v$  channel subtype expressed in shark electrosensory (ampullary) organs (Extended Data Fig. 1e). Furthermore, both skate and shark orthologues contain a charged motif within the S2–S3 region of the IV repeat domain that confers low voltage threshold<sup>6</sup> to skate  $\text{Ca}_v1.3$  (Extended Data Fig. 1g). Shark ampullary organs also expressed several  $\text{Ca}_v$  auxiliary subunits, and the  $I_{\text{CaV}}$  current density and activation threshold exhibited in shark electrosensory cells was similar to that of skates (Extended Data Fig. 1h, i). Our results therefore suggest that  $\text{Ca}_v1.3$  mediates the major depolarizing current in both skate and shark electrosensory cells.

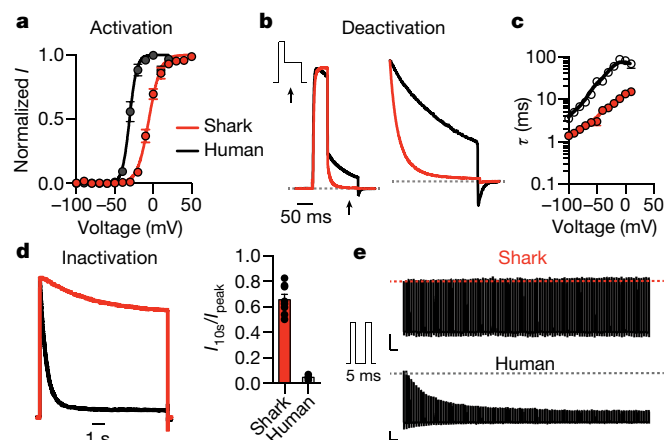
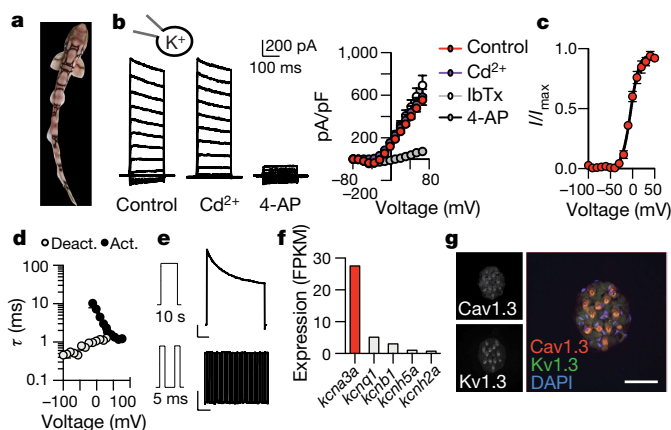
In skate electrosensory cells,  $\text{Ca}^{2+}$  influx activated outward  $\text{K}^+$  currents at relatively negative potentials to occlude inward  $\text{Ca}^{2+}$  currents<sup>6</sup> (Extended Data Fig. 2a–c). In shark cells, however, we observed large  $\text{K}^+$  currents that were activated at more positive voltages and did not affect the amplitude of inward current, which suggests reduced functional interaction between  $\text{Ca}^{2+}$  and  $\text{K}^+$  currents (Extended Data Fig. 2a–c). Indeed,  $\text{K}^+$  currents were not affected by an  $I_{\text{CaV}}$  blocker or a BK

antagonist (Fig. 1b, Extended Data Fig. 2d). Instead, the voltage-gated  $\text{K}_v$  channel inhibitor 4-aminopyridine (4-AP) blocked outward currents from shark electrosensory cells ( $I_{\text{KV}}$ , Fig. 1b).  $I_{\text{KV}}$  was selective for  $\text{K}^+$  and exhibited a relatively high voltage-activation threshold (Fig. 1c, Extended Data Fig. 2e, f). Furthermore, we observed fast activation and deactivation kinetics, whereas voltage-dependent inactivation (similar to desensitization) in response to prolonged voltage pulses was weak and slow, which results in a  $\text{K}^+$  conductance of undiminished current amplitude even after repeated activation–deactivation cycles (Fig. 1d, e, Extended Data Fig. 2g). Among voltage-gated  $\text{K}^+$  channels, transcripts that encode the pore-forming subunit of  $\text{K}_v1.3$  predominated in shark ampullary organs (together with several  $\text{K}_v$  auxiliary subunits) and co-localized with  $\text{Ca}_v1.3$  in electrosensory cells (Fig. 1f, g, Extended Data Fig. 2h). Outside of ampullary organs, only a truncated form of  $\text{K}_v1.3$  that lacks an essential N-terminal tetramerization domain was observed in the brain (Extended Data Fig. 2i).  $\text{K}_v1.3$  expression was not detected in skate ampullary organs, and only shark electrosensory cells exhibited 4-AP-sensitive voltage-gated  $\text{K}^+$  currents (Extended Data Fig. 3a, b). Furthermore, both shark and skate electrosensory cells expressed BK transcripts that, when heterologously expressed, produced channels with similar properties; however, functional BK currents were observed only in the latter (Extended Data Fig. 3). As such, BK channels do not contribute appreciably to the major  $\text{K}^+$  conductance in shark electrosensory cells, at least not under the developmental or physiological conditions examined here. In summary, shark electrosensory cells express a specific  $I_{\text{KV}}$  with voltage-dependent properties that support repetitive stimulation.

Shark  $\text{K}_v1.3$  is 80% identical to the human orthologue, but its voltage threshold for activation was shifted to more depolarized values compared with human  $\text{K}_v1.3$  (Fig. 2a, Extended Data Fig. 4a). Furthermore, shark  $\text{K}_v1.3$  was activated at slightly slower rates and deactivated with rapid kinetics, requiring substantially less negative voltage to return to the resting state compared with the human channel (Fig. 2b, c, Extended Data Fig. 4b, c). Shark  $\text{K}_v1.3$  inactivation was slow and only weakly voltage-dependent compared to human  $\text{K}_v1.3$  (Fig. 2d, Extended Data Fig. 4d). Consequently, shark  $\text{K}_v1.3$  produced a conductance that could be repetitively stimulated with undiminished amplitude, whereas human  $\text{K}_v1.3$  quickly inactivated with repetitive voltage pulses (Fig. 2e). These biophysical properties resemble those of native shark  $I_{\text{KV}}$ , which also exhibited a comparable pharmacological profile (Extended Data Fig. 4e, f). One notable difference is that the deactivation kinetics of native  $I_{\text{KV}}$  were faster than those of the cloned channel, particularly at positive voltages. As such, although  $\text{K}_v1.3$  forms the predominant  $\text{K}^+$  conductance in shark electrosensory cells, additional regulatory mechanisms may be provided by auxiliary subunits, signalling cascades or structural proteins to further enhance rapid deactivation.

The voltage-dependent properties of shark  $\text{K}_v1.3$  probably derive from altered voltage-sensor domain movements, which we verified by comparing gating currents from modified non-conductive shark and human  $\text{K}_v1.3$ . For these experiments, we analysed a human isoform that

<sup>1</sup>Department of Physiology, University of California, San Francisco, San Francisco, CA, USA. <sup>2</sup>Present address: Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA, USA. <sup>3</sup>These authors contributed equally: Nicholas W. Bellono, Duncan B. Leitch. \*e-mail: nbellono@harvard.edu; david.julius@ucsf.edu

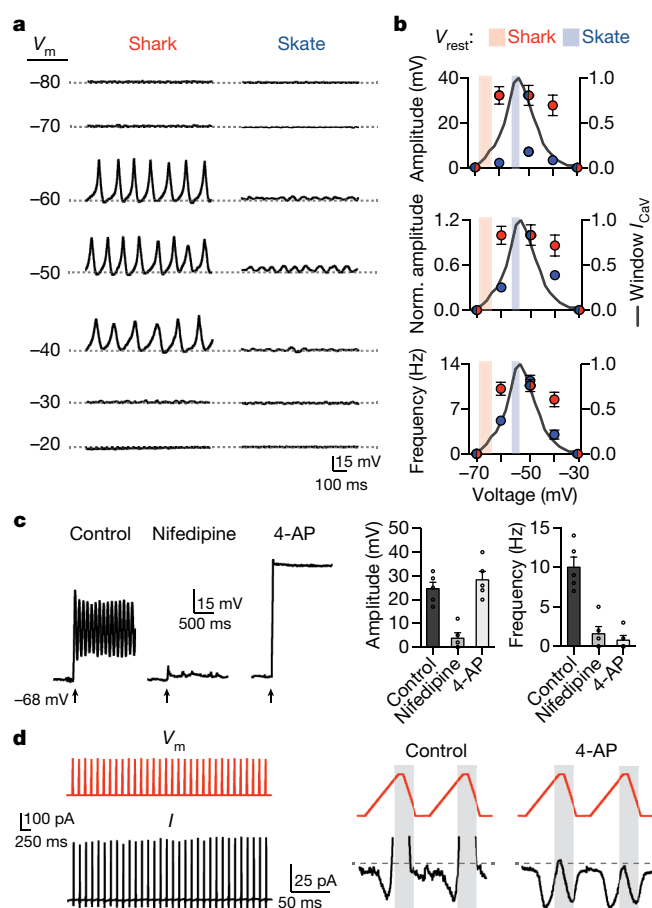


exhibits increased surface expression and therefore enhanced gating current amplitude<sup>8</sup> (Extended Data Fig. 4g). Similar to ion (permeating) currents, upward (activating) voltage sensor movements (represented as ON gating charge,  $Q_{ON}$ ) for shark  $K_V1.3$  exhibited a higher voltage-activation threshold and slower kinetics compared with human  $K_V1.3$  (Extended Data Fig. 5a–c). Moreover, shark  $K_V1.3$  gating-current deactivation ( $Q_{OFF}$ ; return of voltage sensors to a resting state after depolarizing pulses) required less negative voltage and was accelerated (Extended Data Fig. 5d–h). Consequently, ion tail-current deactivation, which represents closure of the channel pore, was faster in shark  $K_V1.3$  than in human  $K_V1.3$  (Extended Data Fig. 5i). We next asked why shark  $K_V1.3$  appears to favour a resting state. When we recorded  $Q_{OFF}$  after a series of depolarizing voltage pulses of varying lengths, human  $K_V1.3$  deactivation kinetics markedly slowed after pulses longer than 1 ms, whereas shark channel deactivation rates remained relatively fast and constant with increasing pulse lengths (Extended Data Fig. 6a–d). This slowing in deactivation is characteristic of voltage-sensor-domain ‘relaxation’, which has been proposed to slow the closure of  $K_V$  channels<sup>9</sup>. We therefore propose that reduced voltage-sensor relaxation in shark  $K_V1.3$  results in decreased stability of the open state; as such, much less negative voltage is required to return channels to a resting state and mediate fast channel closure (depicted in our model, Extended Data Fig. 6e).

We analysed shark–human  $K_V1.3$  chimaeras to see if specific domains specify relevant biophysical attributes. Replacement of the S1–S6 region of human  $K_V1.3$  with that of shark recapitulated the high voltage-activation threshold, rapid deactivation kinetics and weak inactivation of wild-type shark  $K_V1.3$  channels, and the converse chimaera also altered channel properties (Extended Data Fig. 7a, b). We next

substituted just the voltage-sensor-domain region (S1–S4) and found that activation threshold and deactivation kinetics were greatly affected, whereas inactivation remained similar (Extended Data Fig. 7a, b). By contrast, only inactivation was affected by replacing S5–S6 (Extended Data Fig. 7c–e), consistent with a role for the outer pore in C-type inactivation<sup>10</sup>. Therefore, specific structural adaptations in the  $K_V1.3$  transmembrane core specify physiologically relevant biophysical properties.

Membrane voltage ( $V_m$ ) oscillations within ampullae control neurotransmitter release from electrosensory cells onto afferent nerves, thereby shaping signals to the central nervous system<sup>2,3</sup>. In skate electrosensory cells, we found that functional coupling of  $Ca_V1.3$  and BK mediates  $V_m$  oscillations that are tuned to low frequencies, such as those detected by behaving animals<sup>6</sup>. Consistent with our previous results, skate electrosensory cells had a resting  $V_m$  of  $-54$  mV, near the peak of the  $I_{CaV}$  window current, at which cells exhibited spontaneous voltage oscillations (Fig. 3a, b). Injecting current to bring the skate cell  $V_m$  to various potentials modulated oscillatory behaviour, markedly changing both frequency and amplitude across physiological membrane potentials (Fig. 3a, b). By contrast, shark electrosensory cells had an average resting  $V_m$  of  $-66$  mV, which is on the cusp of the  $I_{CaV}$  window current at which cells were relatively quiet (Fig. 3a, b). Injecting current to bring  $V_m$  to within the range of the  $I_{CaV}$  threshold and window current elicited robust, repetitive  $V_m$  spiking that lasted for the duration of the recordings. Spiking only slightly decreased in amplitude and frequency at more positive voltages (Fig. 3a, b), because spike amplitude is probably determined by the voltage threshold of  $I_{KV}$ , in addition to  $I_{CaV}$  window current. Indeed, the  $I_{CaV}$  inhibitor nifedipine blocked evoked depolarization and 4-AP prevented or greatly slowed repolarization (Fig. 3c). As such, in both shark and skate, activity is limited to membrane voltages in which  $I_{CaV}$  window current is observed, but the dynamics in the two species are markedly different: the system behaves as an all-or-none ON/OFF switch in sharks, but it is more tunable in skates.



**Fig. 3 | Voltage dynamics in electrosensory cells.** **a**,  $V_m$ -dependent spiking in shark electrosensory cells and smaller voltage oscillations from skate cells at the indicated membrane potentials (dotted lines) achieved by current injection. **b**, Average voltage oscillation amplitude, normalized amplitude, and frequency at various membrane voltages for shark (red) and skate (blue) electrosensory cells. Values for oscillations from skate cells were significantly different for voltages at which activity was observed, whereas shark-cell activity changed only slightly at more depolarized voltages.  $n = 5$ ,  $P < 0.01$  for skate at -40, -50, -60 mV versus all other voltages, two-way ANOVA with post hoc Tukey test. The mean  $\pm$  s.e.m. for resting membrane voltages ( $V_{rest}$ ) for shark (red,  $-66.2 \pm 2.7$  mV) and skate (blue,  $-54 \pm 1.8$  mV) are indicated as bars on each graph.  $n = 5$  for skate and 10 for shark,  $P < 0.01$  two-tailed Student's  $t$ -test. **c**, Brief current injection (2 pA, 5 ms, at arrow) from -68 mV elicited repetitive spiking that was inhibited by nifedipine. Brief injection elicited sustained depolarization in the presence of 4-AP.  $n = 5$ ,  $P < 0.001$  for amplitude of control versus nifedipine and for control versus nifedipine or 4-AP for frequency, one-way ANOVA. **d**, Left, simulated repetitive spiking voltage-clamp protocol ( $V_m$ , -65 mV to -15 mV) elicited currents that do not inactivate ( $I$ ); right, depolarizing voltage increased inward currents until a voltage threshold was reached that activated a large outward current (shaded grey area), which rapidly deactivated with hyperpolarizing voltage. The cycle repeats in the next simulated spike protocol. The outward current was blocked by 4-AP, revealing a resurgent inward current in the hyperpolarizing phase. Representative of  $n = 4$ . All data are represented as mean  $\pm$  s.e.m.,  $n$  denotes cells.

To further assess how the kinetics and voltage dependence of  $I_{CaV}$  and  $I_{KV}$  contribute to  $V_m$  spiking dynamics, we recorded currents in response to simulated  $V_m$  spikes in the voltage-clamp mode. Repetitive simulated spikes induced inward and subsequent outward currents that rapidly activated and deactivated but did not inactivate (Fig. 3d). Inward  $I_{CaV}$  was small at a resting potential of -65 mV and increased with positive voltage until a high voltage threshold, which triggered an outward current that rapidly deactivated when the voltage was returned to resting levels (Fig. 3d). 4-AP blocked this outward current—which

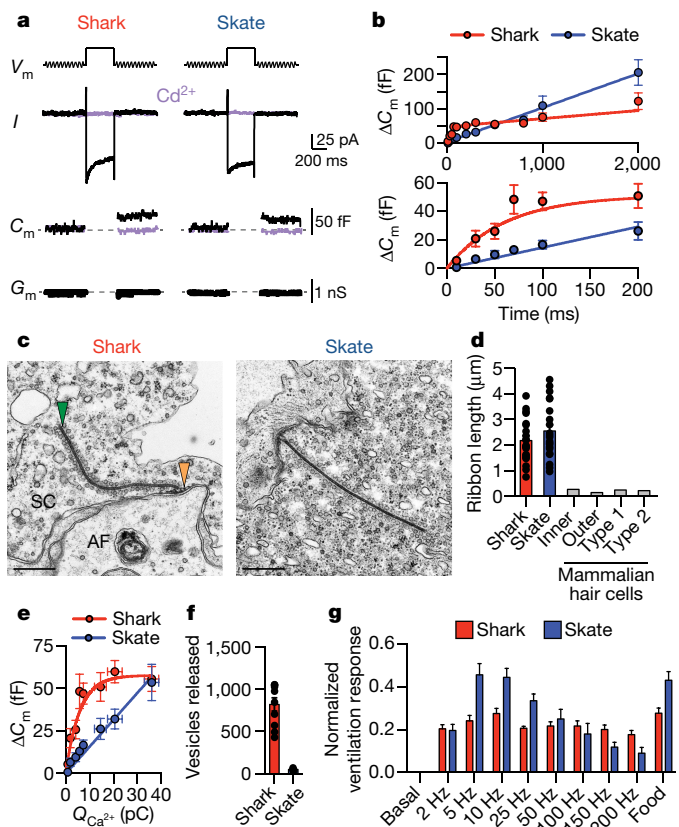
implicates  $I_{KV}$ —while also revealing a resurgent  $I_{CaV}$  upon hyperpolarization that could contribute to the initiation of the following voltage spike (Fig. 3d). Therefore, the low voltage threshold and inactivation properties of  $I_{CaV}$ , coupled with the high voltage threshold, rapid deactivation and weak inactivation of  $I_{KV}$ , are suited to cooperatively mediate  $V_m$  spiking in shark electrosensory cells. Notably, transcripts for  $Ca^{2+}$ -binding proteins and pumps were enriched in shark ampullary organs, which may help to facilitate repetitive  $Ca^{2+}$  influx (Extended Data Fig. 8a). Together, these cellular properties could support robust repetitive activity in shark electrosensory cells to amplify responses to incoming electrical signals. By contrast, skate cells exhibit low-level tonic activity at rest that could be retimed or modulated in response to particular incoming electrical frequencies to alter neurotransmitter release.

To determine how synaptic vesicle release is affected by differences in  $V_m$  activity, we monitored membrane capacitance ( $C_m$ ) to measure vesicle fusion in response to electrical stimuli. Depolarization of shark or skate electrosensory cells elicited inward currents and capacitance changes that were blocked by  $Cd^{2+}$ , which indicates that  $I_{CaV}$  is required for vesicular release (Fig. 4a). Increasing the stimulus duration increased changes in  $C_m$  (Fig. 4b, Extended Data Fig. 9a), but with distinct dynamics in shark compared with skate electrosensory cells: brief voltage stimuli induced larger changes in shark-cell  $C_m$  that saturated in response to longer stimuli, whereas skate cells exhibited  $C_m$  changes that increased linearly with the duration of the stimulus (Fig. 4b).

Differences in the number or distribution of synaptic vesicles could account for these distinct vesicle release dynamics. Skate electrosensory cells contain large synaptic ribbons that tether numerous vesicles for release onto postsynaptic afferents<sup>11,12</sup>. Indeed, shark and skate ampullae expressed transcripts associated with ribbon synapses, and ultrastructural analysis of electrosensory cells revealed that both contain remarkably long synaptic ribbons compared to those from mammalian hair cells<sup>13</sup> (Fig. 4c, d, Extended Data Fig. 8b, c). Moreover, shark and skate ribbons were similarly shaped and tethered an equivalent number of vesicles of comparable diameter. However, shark electrosensory cells had a larger 'readily releasable' vesicle pool (Extended Data Fig. 8d), which facilitates rapid and efficient exocytosis in other systems<sup>14,15</sup> and might account for our observation that short voltage stimuli induce larger changes in  $C_m$  in electrosensory cells from the shark. Conversely, skate electrosensory cells contained more free cytosolic vesicles that, by analogy with other systems<sup>14</sup>, may represent a larger 'refilling pool' of recently generated vesicles available for tethering and release (Extended Data Fig. 8d). Thus, skate cells may be better equipped to continuously supply vesicles for tonic release, reflected by the non-saturating, linear change in  $C_m$  in response to increasing the voltage of the stimulus.

To determine the number of vesicles released in response to a single  $V_m$  spike or oscillation, we first integrated  $I_{CaV}$  ( $Q_{Ca^{2+}}$ ) elicited by stimuli of varying duration, thereby establishing a relationship between  $Q_{Ca^{2+}}$  and  $C_m$  (Fig. 4e). Shark and skate cells responded equally to identical voltage stimuli, further suggesting that  $I_{CaV}$  is similar in both cell types and that  $K^+$ -channel identity dictates the oscillation phenotype and resulting amplitude of  $I_{CaV}$  (Extended Data Fig. 9b–d). We next fit  $Q_{Ca^{2+}}$  induced by simulated  $V_m$  spikes or oscillations to  $Q_{Ca^{2+}}-C_m$  relationships, and used the specific capacitance for a pure lipid membrane with diameter equal to that of an electrosensory cell vesicle to estimate fusion events from measured changes in  $C_m$ <sup>16</sup>. Notably, our calculations suggest that shark cells released at least ten times more vesicles in response to one  $V_m$  spike compared to skate cells subjected to a single oscillation event (Fig. 4f). Furthermore, as with other ribbon synapses, the large storage pool could facilitate a sustained release of vesicles in response to repetitive  $V_m$  spiking, further amplifying the signals<sup>14</sup>. Taken together, these properties should render sharks acutely sensitive to incoming signals by greatly amplifying vesicular release to even very brief stimuli. By contrast, skate cells may better encode stimulus variation with tunable voltage oscillations and more graded vesicle release.

We next asked how differences in cellular dynamics might contribute to sensation at the organismal level. Elasmobranchs preferentially



**Fig. 4 | Tuning of electrosensory cell vesicular release and electrosensation.** **a**, Representative capacitance measurements. A sine wave was applied before and after a 200-ms,  $-20$ -mV pulse to activate  $I_{CaV}$  in shark (red) or skate (blue) electrosensory cells ( $V_m$ ).  $I_{CaV}$  ( $I$ ) and capacitance ( $C_m$ ) changes were blocked by  $Cd^{2+}$  (purple). Membrane conductance ( $G_m$ ) was constant in all quantified data. Representative of  $n = 4$ . **b**, Top, average capacitance changes in response to various durations of voltage stimuli. The response relationship from shark electrosensory cells (red) saturated in response to prolonged stimuli, whereas capacitance changes in skate cells (blue) increased in a nearly linear fashion with increasing duration of the voltage stimulus. Bottom, brief stimuli elicited larger capacitance changes in shark electrosensory cells compared with skate cells. The relationship between capacitance and duration of the voltage stimulus in shark cells was best fit by an exponential relationship that plateaued at 51 fF with a time constant of 42 ms, whereas the relationship in skate cells was linear with a slope of  $0.15 \pm 0.01$  fF  $ms^{-1}$ .  $n = 6$ . **c**, Micrographs showing shark and skate ribbon synapses. The electrosensory cell (SC), afferent nerve (AF), synapse (orange arrowhead) and synaptic ribbon (green arrowhead) are indicated. Scale bars, 500 nm. **d**, Average ribbon length was similar in shark (red) and skate (blue) electrosensory cells, and is much longer than those from homologous mammalian hair cells<sup>13</sup>.  $n = 21$ . **e**, Average capacitance change elicited by integrated  $I_{CaV}$  ( $Q_{Ca^{2+}}$ ) revealed that less  $Q_{Ca^{2+}}$  is required to elicit large capacitance changes in shark electrosensory cells (red) compared with skate cells (blue). The relationship between capacitance and  $Q_{Ca^{2+}}$  was exponential for shark and linear for skate.  $n = 6$ . **f**, Average number of calculated vesicles released per spike or oscillation based on the relationship between capacitance and  $Q_{Ca^{2+}}$  and simulated voltage oscillation-induced  $Q_{Ca^{2+}}$ .  $n = 6$ ,  $P < 0.0001$ , two-tailed Student's  $t$ -test. **g**, Normalized ventilatory responses elicited from live sharks or skates in response to 50- $\mu$ V electric stimuli at indicated frequencies. Food odors were used as a positive control. Shark responses to electrical frequencies were largely comparable, regardless of stimulus frequency. Skate peak responses to low frequencies were significantly different from those at other frequencies but were comparable to odorant-elicited responses ( $n = 10$ ,  $P < 0.0001$  for 5, 10, 25 Hz versus other frequencies, one-way ANOVA with multiple comparisons). All data are represented as mean  $\pm$  s.e.m,  $n$  denotes cells, synaptic ribbons, or behavioural trials.

respond to low frequency electrical signals produced by their prey, but direct comparison of frequency selectivity between species is confounded by behavioural and physiological differences<sup>5</sup>. We therefore measured changes in ventilatory rate as a basic, time-locked physiologic metric that is well-validated in electroreceptive elasmobranch species and readily observed in response to sensory cues, such as weak electrical stimuli or odors<sup>17,18</sup> (Supplementary Videos 1 and 2). When presented with electrical stimuli of identical strength, the ventilation rates of sharks increased similarly at all stimulus frequencies (Fig. 4g). By contrast, the ventilation rate of skates maximally increased at low frequencies, resembling voltage oscillations in their electrosensory cells and signals emitted by their electric organ (Fig. 4g). In both species, maximal electrically induced ventilatory responses were similar to those evoked by food odors as a comparative control (Fig. 4g). As such, shark electroreception may act as a threshold detector for broad frequencies, potentially reflecting its role in predation. By contrast, skate sensation appears more specifically tuned to enable the detection of signals from prey as well as frequencies in the range of conspecific electric-organ discharges<sup>5</sup>.

Electroreception has independently evolved in many taxa to facilitate particular behaviours ranging from predation to communication. By analysing related species that use electroreception for distinct purposes, we found that subtle molecular variations considerably alter cellular properties that could ultimately mediate differences in behaviour. Our results suggest that molecular tuning of  $V_m$  oscillations in electrosensory cells is important for the initial detection and discrimination of salient electrical signals, although anatomical characteristics and processing by the central nervous system probably contribute to additional signal filtering<sup>19</sup> (Extended Data Fig. 10). This observation is reminiscent of other sensory modalities in which sensory cells or their receptors are modified to mediate the detection of relevant stimuli. For example, expression and regulation of ion channels enable hair cells—which are developmentally related to electrosensory cells—to produce  $V_m$  oscillations of various amplitudes and frequencies to mediate detection of particular auditory signals<sup>20,21</sup>. Although it is not known how broadly  $V_m$  oscillation tuning applies to electroreception, the electrosensory organs of paddlefish also express transcripts for  $Ca_v$  and  $K^+$  channels, which suggests that similar mechanisms may produce  $V_m$  oscillations in these systems<sup>22,23</sup>. Furthermore, weakly electric fish use oscillating or spiking electrosensory organs to facilitate conspecific communication<sup>24</sup>. Our results demonstrate one mechanism by which  $K^+$ -channel modification shapes electrosensory cell activity, but differential regulation of  $K^+$  channels or other transduction components could provide alternative tuning mechanisms in distinct species or under specific developmental or physiological states to facilitate dynamic electroreceptive behaviours<sup>5,25,26</sup>.

## Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0160-9>.

Received: 6 October 2017; Accepted: 4 April 2018;  
Published online: 30 May 2018

- Josberger, E. E. et al. Proton conductivity in ampullae of Lorenzini jelly. *Sci. Adv.* **2**, e1600112 (2016).
- Clusin, W. T. & Bennett, M. V. The oscillatory responses of skate electroreceptors to small voltage stimuli. *J. Gen. Physiol.* **73**, 685–702 (1979).
- Teeter, J. H. & Bennett, M. V. L. Synaptic transmission in the ampullary electroreceptor of the transparent catfish, *Kryptopterus*. *J. Comp. Physiol.* **142**, 371–377 (1981).
- Kalmijn, A. J. The electric sense of sharks and rays. *J. Exp. Biol.* **55**, 371–383 (1971).
- Sisneros, J. A. & Tricas, T. C. Neuroethology and life history adaptations of the elasmobranch electric sense. *J. Physiol. Paris* **96**, 379–389 (2002).
- Bellono, N. W., Leitch, D. B. & Julius, D. Molecular basis of ancestral vertebrate electroreception. *Nature* **543**, 391–396 (2017).
- King, B. L., Shi, L. F., Kao, P. & Clusin, W. T. Calcium activated  $K^+$  channels in the electroreceptor of the skate confirmed by cloning. Details of subunits and splicing. *Gene* **578**, 63–73 (2016).

8. Kubota, T., Correa, A. M. & Bezanilla, F. Mechanism of functional interaction between potassium channel Kv1.3 and sodium channel NavBeta1 subunit. *Sci. Rep.* **7**, 45310 (2017).
9. Labro, A. J., Lacroix, J. J., Villalba-Galea, C. A., Snyders, D. J. & Bezanilla, F. Molecular mechanism for depolarization-induced modulation of Kv channel closure. *J. Gen. Physiol.* **140**, 481–493 (2012).
10. Hoshi, T. & Armstrong, C. M. C-type inactivation of voltage-gated K<sup>+</sup> channels: pore constriction or dilation? *J. Gen. Physiol.* **141**, 151–160 (2013).
11. Sejnowski, T. J. & Yodkowski, M. L. A freeze-fracture study of the skate electroreceptor. *J. Neurocytol.* **11**, 897–912 (1982).
12. Fields, R. D. & Ellisman, M. H. Synaptic morphology and differences in sensitivity. *Science* **228**, 197–199 (1985).
13. Nouvian, R., Beutner, D., Parsons, T. D. & Moser, T. Structure and function of the hair cell ribbon synapse. *J. Membr. Biol.* **209**, 153–165 (2006).
14. Matthews, G. & Fuchs, P. The diverse roles of ribbon synapses in sensory neurotransmission. *Nat. Rev. Neurosci.* **11**, 812–822 (2010).
15. Graydon, C. W., Cho, S., Li, G. L., Kachar, B. & von Gersdorff, H. Sharp Ca<sup>2+</sup> nanodomains beneath the ribbon promote highly synchronous multivesicular release at hair cell synapses. *J. Neurosci.* **31**, 16637–16650 (2011).
16. Hille, B. *Ion channels of excitable membranes* 3rd edn (Sinauer, Sunderland, 2001).
17. Peters, R. C. & Evers, H. P. Frequency selectivity in the ampullary system of an elasmobranch fish (*Scyliorhinus canicula*). *J. Exp. Biol.* **118**, 99–109 (1985).
18. Kempster, R. M., Hart, N. S. & Collin, S. P. Survival of the stillest: predator avoidance in shark embryos. *PLoS ONE* **8**, e52551 (2013).
19. Montgomery, J. C. & Bodznick, D. Signals and noise in the elasmobranch electrosensory system. *J. Exp. Biol.* **202**, 1349–1355 (1999).
20. Fettiplace, R. & Fuchs, P. A. Mechanisms of hair cell tuning. *Annu. Rev. Physiol.* **61**, 809–834 (1999).
21. Rutherford, M. A. & Roberts, W. M. Spikes and membrane potential oscillations in hair cells generate periodic afferent activity in the frog sacculus. *J. Neurosci.* **29**, 10025–10037 (2009).
22. Modrell, M. S. et al. Insights into electrosensory organ development, physiology and evolution from a lateral line-enriched transcriptome. *eLife* **6**, e24197 (2017).
23. Neiman, A. B. & Russell, D. F. Two distinct types of noisy oscillators in electroreceptors of paddlefish. *J. Neurophysiol.* **92**, 492–509 (2004).
24. Baker, C. A., Huck, K. R. & Carlson, B. A. Peripheral sensory coding through oscillatory synchrony in weakly electric fish. *eLife* **4**, e08163 (2015).
25. Meyer, J. H. & Zakon, H. H. Androgens alter the tuning of electroreceptors. *Science* **217**, 635–637 (1982).
26. Sisneros, J. A., Tricas, T. C. & Luer, C. A. Response properties and biological function of the skate electrosensory system during ontogeny. *J. Comp. Physiol. A* **183**, 87–99 (1998).

**Acknowledgements** We thank S. Bennett from the Marine Biological Laboratory for supplying animals, J. Wong from the Gladstone/University of California, San Francisco (UCSF) transmission electron microscopy core for performing electron microscopy, A. Zimmerman for help with capacitance measurements, discussion and reading of the manuscript, R. Edwards for discussion, and R. Nicoll for input on the manuscript. This work was supported by a National Institutes of Health (NIH) Institutional Research Service Award to the UCSF CVRI (T32HL007731 to N.W.B.), a Howard Hughes Medical Institute Fellowship of the Life Sciences Research Foundation (N.W.B.), a Simons Foundation Postdoctoral Fellowship to the Jane Coffin Childs Memorial Fund (D.B.L.) and grants from the NIH (1K99DC016658 to D.B.L., K99DK115879 to N.W.B., and NS055299 and NS105038 to D.J.).

**Reviewer information** *Nature* thanks B. Carlson, C. Lingle, H. von Gersdorff and the other anonymous reviewer(s) for their contribution to the peer review of this work.

**Author contributions** N.W.B. designed and performed electrophysiological studies, D.B.L. designed and performed gene expression, anatomical and behavioural studies, and N.W.B., D.B.L. and D.J. wrote the manuscript.

**Competing interests** The authors declare no competing interests.

#### Additional information

**Extended data** is available for this paper at <https://doi.org/10.1038/s41586-018-0160-9>.

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41586-018-0160-9>.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>.

**Correspondence and requests for materials** should be addressed to N.W.B. or D.J.

**Publisher's note**: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## METHODS

**Animals and cells.** Male and female chain catsharks (*Scyliorhinus retifer*) and little skates (*Leucoraja erinacea*) were provided by the Marine Biological Laboratory (Woods Hole, MA, USA) and their use was approved by the UCSF Animal Care and Use Committee. Animals were euthanized with tricaine methanesulfonate (MS222,  $1 \text{ g l}^{-1}$ ). Ampullary organs of adult animals were removed from the hyoid cluster (skates) or buccal and supraorbital clusters (sharks) on ice and further dissected by removing most of the canals and afferent nerve fibres. Ampullae were treated with papain for less than 5 min and then electrosensory cells were mechanically dissociated over the recording chamber. Isolated electrosensory cells were identified by the presence of their single kinocilium. HEK293T cells (American Type Culture Collection, ATCC) were grown in DMEM, 10% fetal calf serum and 1% penicillin/streptomycin at  $37^\circ\text{C}$  in 5%  $\text{CO}_2$ . Cells were transfected with Lipofectamine 2000 (Invitrogen) according to the manufacturer's protocol. Cell lines were verified from the ATCC but were not further tested for identity or mycoplasma contamination during our studies. 100 ng of  $\text{Kv}1.3$  or  $1 \mu\text{g}$  of non-conducting  $\text{Kv}1.3$  or BK constructs were co-expressed with  $0.3 \mu\text{g}$  GFP. Mock transfection experiments ( $0.3 \mu\text{g}$  GFP) were performed as controls, in which minimal voltage-activated outward current was observed.

**Whole mount preparations.** Juvenile fish were euthanized with an overdose of MS-222 in artificial seawater and fixed in 4% paraformaldehyde for at least 24 h. The cartilage matrix and electroreceptor tubules were stained using Alcian Blue (20 mg Alcian Blue 8GX in 30 ml glacial acetic acid and 70 ml 100% ethanol) and bone was stained using Alizarin Red following previously published methods<sup>27</sup>.

**Molecular biology.** *kcn3a* and *kcnma1a* from shark ampullary organs were synthesized by Genscript. Human *KCNA3* was from Genscript, skate BK was from the Julius laboratory, and mouse *Kcnma1* was a gift from L. Salkoff (Addgene plasmid 16195). Chimaera synthesis and mutagenesis were carried out and verified by Genscript or by the QuikChange Lightning site-directed mutagenesis kit (Agilent Genomics).

**Electrophysiology.** Recordings were carried out at room temperature using a MultiClamp 700B amplifier (Axon Instruments) and digitized using a Digidata 1322A (Axon Instruments) interface and pClamp software (Axon Instruments). Capacitance and associated ion current measurements were amplified and digitized with an EPC10 amplifier in lock-in mode (HEKA) and Patchmaster software (HEKA). Unless stated otherwise, whole-cell data were filtered at 1 kHz and sampled at 10 kHz, and single-channel data were filtered at 5 kHz and sampled at 50 kHz. Data were leak subtracted online using a P/4 protocol, except for data obtained using voltage ramp protocols, and membrane potentials were corrected for liquid junction potentials. Electrosensory cell recordings were performed using borosilicate glass pipettes polished to 8–10 M $\Omega$ . For heterologous expression experiments in HEK293, recordings were performed using pipettes polished to 2–3 M $\Omega$ .

The extracellular solution was a modified 'elasmobranch Ringer's solution' containing (in mM): 250 NaCl, 6 KCl, 4  $\text{CaCl}_2$ , 1  $\text{MgCl}_2$ , 10 glucose, 5 HEPES, 360 urea, pH 7.6. When analysing native  $\text{K}^+$  current properties,  $500 \mu\text{M}$   $\text{Cd}^{2+}$  was included to block  $I_{\text{CaV}}$ . Four intracellular solutions were used, as follows: for recording  $I_{\text{CaV}}$  (in mM): 250 CsMeSO<sub>3</sub>, 1  $\text{MgCl}_2$ , 11 Cs-EGTA, 10 HEPES, 30 sucrose, 360 urea, pH 7.6; for recording  $I_{\text{K}}$  (in mM): 250 K-gluconate, 1  $\text{MgCl}_2$ , 11 K-EGTA, 10 HEPES, 30 sucrose, 360 urea, pH 7.6; for recording membrane potential (in mM): 250 K-gluconate, 1  $\text{MgCl}_2$ , 1 K-EGTA, 10 HEPES, 20 sucrose, 2 MgATP, 360 urea, pH 7.6; for recording capacitance changes (in mM): 250 CsMeSO<sub>3</sub>, 1  $\text{MgCl}_2$ , 1 Cs-EGTA, 2 MgATP, 10 HEPES, 20 sucrose, 360 urea, pH 7.6. For recording heterologous  $\text{Kv}1.3$  ionic current, the intracellular solution contained (in mM): 145 K-gluconate, 5 KCl, 1  $\text{MgCl}_2$ , 5 K-EGTA, 10 HEPES, 10 sucrose, pH 7.2. The extracellular solution contained (in mM): 150 NaCl, 5 KCl, 2  $\text{CaCl}_2$ , 2  $\text{MgCl}_2$ , 10 HEPES, 10 glucose, pH 7.4. For measuring  $\text{Kv}1.3$  gating currents, the intracellular solution contained (in mM): 150 NMDGMeSO<sub>3</sub>, 1  $\text{MgCl}_2$ , 10 Cs-EGTA, 10 HEPES, 10 sucrose, pH 7.3. The extracellular solution contained (in mM): 150 TEACl, 1  $\text{CaCl}_2$ , 1  $\text{MgCl}_2$ , 10 HEPES, 10 glucose, pH 7.3. For gating current measurements, we used non-conducting shark (W407F) and human (W436F)  $\text{Kv}1.3$  channels<sup>28,29</sup>. A short isoform of human  $\text{Kv}1.3$ , which exhibits identical gating properties but increased surface expression<sup>8</sup>, was used to increase the amplitude of the gating current, whereas the long isoform was used for most ion current studies. For BK single-channel recordings, the intracellular solution contained (in mM): 136 K-gluconate, 4 KCl, 1 K-EGTA, 1 HEDTA, 10 HEPES, 10 glucose, pH 7.3. The extracellular solution contained (in mM): 136 K-gluconate, 4 KCl, 1  $\text{MgCl}_2$ , 10 HEPES, 10 glucose, pH 7.3. Calculated concentrations of buffered  $\text{Ca}^{2+}$  added to intracellular solution were made using MaxChelator (C. Patton, Stanford University).

The pharmacological inhibitors or agonists Bay K, nifedipine, mibefradil, 4-AP, XE991, NS11021 and quinidine were from Tocris. Iberiotoxin,  $\alpha$ -dendrotoxin, margatoxin, UK78282 and Guangxitoxin-1E were from Alamone Labs. Compounds were dissolved in <1% vehicle (DMSO or water), which was used as a control. Ionic pore-blocker stocks were prepared in standard extracellular

solution and diluted before use. Unless stated otherwise, the following concentrations were used: 100 or 500  $\mu\text{M}$   $\text{Cd}^{2+}$ , 1  $\mu\text{M}$  Bay K, 10  $\mu\text{M}$  nifedipine, 5  $\mu\text{M}$  mibefradil, 1 mM 4-AP, 100 nM iberiotoxin, 10 mM TEA<sup>+</sup>, 10  $\mu\text{M}$  NS11021, 100  $\mu\text{M}$  quinidine, 25 nM  $\alpha$ -dendrotoxin, 10 nM margatoxin, 1  $\mu\text{M}$  UK78282, 25 nM Guangxitoxin-1E, 20  $\mu\text{M}$  XE991. Pharmacological effects were quantified by differences in normalized peak current from the same cell after bath application of the drug ( $I_{\text{treatment}}/I_{\text{control}}$ ).

Unless stated otherwise,  $I_{\text{CaV}}$  was measured in response to 200-ms voltage pulses in 10-mV increments from a  $-100 \text{ mV}$  holding potential.  $G$ - $V$  relationships were derived from  $I$ - $V$  curves by calculating  $G$  according to the following:  $G = I_{\text{Ca}}/(V_{\text{m}} - E_{\text{rev}})$ , and fit with a Boltzmann equation. Voltage-dependent inactivation was measured during  $-20 \text{ mV}$  voltage pulses after a series of 2-s prepulses ranging from  $-100$  to  $60 \text{ mV}$  in 10-mV increments. Voltage-dependent inactivation was quantified as  $I/I_{\text{max}}$ , with  $I_{\text{max}}$  occurring at the activating voltage pulse after a  $-100 \text{ mV}$  prepulse.  $I_{\text{K}}$  was measured in response to 200-ms voltage pulses in 10-mV increments from a  $-100 \text{ mV}$  holding potential.  $G$ - $V$  relationships were established from normalized tail currents measured at  $-30 \text{ mV}$  after 50-ms voltage pulses in 10-mV increments from a  $-100 \text{ mV}$  holding potential. Voltage-dependent inactivation was measured during 40-mV voltage pulses after a series of 5-s prepulses ranging from  $-100$  to  $60 \text{ mV}$  in 10-mV increments. Cumulative inactivation was measured in response to 50-ms, 40-mV pulses every 5 ms. Activation kinetics were determined by fitting the initial rising phase of currents activated by various voltages with single exponentials. Deactivation kinetics were fit with a single exponential upon repolarization to various voltages in 10-mV increments from  $40 \text{ mV}$  to  $-100 \text{ mV}$  after 40-mV prepulses of the indicated durations. The reversal potential for  $I_{\text{KV}}$  was measured by plotting tail current-voltage relationships using a similar protocol that stepped in 10-mV increments from  $40 \text{ mV}$  to  $-120 \text{ mV}$ . Single-channel currents were measured from the middle of the noise band between closed and open states or calculated from the difference between Gaussian-fitted closed and open peaks on all-points amplitude histograms for each excised patch record. Conductance was calculated from the linear slope of  $I$ - $V$  relationships. In current-clamp mode, current injection was used to bring membrane potential to various values and then was fixed. In this case, membrane potential was defined as the base of the spiking or oscillating activity. Alternatively, brief current injection was delivered to determine the effects of pharmacological inhibitors on spiking activity.

$Q_{\text{ON}}$  and  $Q_{\text{OFF}}$  represents the integral of non-linear gating current measured during and after voltage pulses from holding potentials of  $-100$  or  $0 \text{ mV}$ .  $Q_{\text{ON}}$  was quantified only from cells with no ionic current. ON gating-current kinetics were quantified by single exponential fits of the slope of decreasing outward current elicited by voltage pulses in 10-mV increments from a  $-100 \text{ mV}$  holding potential. OFF gating-current kinetics were calculated by single-exponential fits of the slope of increasing negative current elicited by voltage pulses in 10-mV increments from a  $0 \text{ mV}$  holding potential. Voltage dependence of deactivation was also assessed by single-exponential fits of currents upon repolarization to various voltages in 10-mV increments from  $20 \text{ mV}$  to  $-100 \text{ mV}$  after 40-mV prepulses of the indicated durations. Deactivation of gating currents was also measured at  $-100 \text{ mV}$  with exponential fits after a series of 40-mV voltage pulses of varying duration from  $0.5 \text{ ms}$  to  $30 \text{ ms}$ .

Capacitance measurements were performed using a 15-mV, 1.5-kHz sinusoidal stimulation protocol applied from  $-90 \text{ mV}$  before and after depolarization pulses of various lengths to acquire pre- and post-stimulus capacitance values. Cells were discarded when the series resistance ( $R_{\text{s}}$ ) changed and exceeded the membrane resistance ( $R_{\text{m}}$ ) or if membrane conductance ( $G_{\text{m}}$ ) varied greatly after depolarizing voltage pulses. Whole-cell ion currents were filtered at 1 kHz and sampled at 10 kHz. Gating currents and capacitance measurements were filtered at 1 kHz and sampled at 20 kHz. Capacitance records were filtered at 100 Hz during offline analyses. Changes in capacitance were measured by averaging capacitance over 200 ms after the depolarizing voltage step and subtracting the averaged capacitance before depolarization. Intracellular ATP was included in all experiments and 100 nM iberiotoxin, 1 mM 4-AP and intracellular  $\text{Cs}^+$  were used to block  $\text{K}^+$  currents. The integral of  $I_{\text{CaV}}$  ( $Q_{\text{Ca}^{2+}}$ ) was used to account for variability in the kinetics of  $I_{\text{CaV}}$ . To calculate vesicle release, we fit  $Q_{\text{Ca}^{2+}}$  induced by simulated spike- or oscillation-voltage protocols to  $Q_{\text{Ca}^{2+}}-C_{\text{m}}$  relationships. When identical voltage protocols were used to elicit  $I_{\text{CaV}}$ ,  $Q_{\text{Ca}^{2+}}$  was the same from skate and shark electrosensory cells, consistent with the similar  $I_{\text{CaV}}$  in these cells.

**Transcriptional sequencing and analysis.** Poly A<sup>+</sup> RNA was extracted from ampullary electrosensory cells, non-electroreceptor covered skin, muscle and forebrain of an adult chain catshark then was reverse-transcribed using SuperScript III kit (Invitrogen). Sequencing libraries were prepared using the Illumina TruSeq Stranded mRNA Library Prep Kit according to the manufacturer's instructions. Libraries were sequenced on the Illumina Hi-Seq 4000 (V. C. Genomics Sequencing Laboratory, University of California, Berkeley) using 150 cycles of paired end reads, producing between 30 million and 40 million inserts for each sample.

Transcriptomes for each sample were assembled de novo using the Trinity suite (version 2.1.0). Sequences were aligned to the zebrafish protein database (NCBI assembly GRCz10) using the blastx tool from NCBI blast (version 2.2.31) using a maximum *E* value of  $1 \times 10^{-5}$ . Reciprocal blastx alignments (using zebrafish protein sequences that aligned to catshark sequences) were performed to the human protein database. Estimates of relative abundance for differential expression comparisons were performed using the RSEM software package within Trinity. These values are reported as FPKM.

**In situ hybridization histochemistry.** Adult chain catsharks were euthanized with an overdose of MS-222 in artificial seawater and transcardially perfused with PBS followed by 4% paraformaldehyde. Ampullary organs were dissected from the buccal and supraorbital clusters and cryo-protected in 30% sucrose in PBS overnight. Cryostat sections (15- $\mu$ m thick) were probed with digoxigenin-labelled cRNA for shark Cav1.3 and fluorescein-labelled cRNA for shark BK and Kv1.3 receptors. Probes were generated by T7/T3 in vitro transcription reactions using a 500-nucleotide fragment of Cav1.3 cDNA (nucleotides 3700 to 4200), a 325-nucleotide fragment of BK cDNA (nucleotides 636 to 961) and a 470-nucleotide fragment of Kv1.3 cDNA (nucleotides 670 to 1140). Hybridization was developed using anti-digoxigenin and anti-fluorescein Fab fragments, followed by incubation with Fast Red and streptavidin-conjugated Dylight 488 (to probe for BK) according to published methods<sup>30</sup>. After hybridization and detection, sections were covered with a coverslip and co-stained with DAPI as a nuclear marker (Prolong Gold Antifade Mountant with DAPI; Invitrogen).

**Transmission electron microscopy.** Tissue samples from the catshark and skate hyoid capsule with electrosensory cells were fixed in 2% glutaraldehyde, 1% paraformaldehyde in 0.1 M sodium cacodylate buffer at pH 7.4, postfixed in 2% osmium tetroxide in the same buffer, stained en bloc with 2% aqueous uranyl acetate, dehydrated in acetone, infiltrated and embedded in LX-112 resin (Ladd Research Industries). Semi-thin sections stained with toluidine blue were prepared to orient and locate the area of interest. Samples were ultrathin sectioned (typically 100 nm) on a Reichert Ultracut S ultramicrotome and counter-stained with 0.8% lead citrate. Grids were examined on a JEOL JEM-1230 transmission electron microscope (JEOL USA, Inc.) and imaged with the Gatan Ultrascan 1000 digital camera (Gatan Inc.). All measurements were performed in Image J (NIH) on electron micrographs adjusted for brightness and contrast (Photoshop CS6, Adobe Systems). Measurements of vesicles followed published methods<sup>31</sup>, with populations attached to the ribbon structure ('attached'), in proximity to the synapse ('readily releasable') and freely filling cytosolic space ('refilling').

To measure ribbon-shape variation, the difference between the traced distance of the ribbon and the distance of a line drawn from the start to the end of the ribbon was divided by the distance of that line. All values represent a positive difference (increase in length) from this straight line. The angle of the ribbon to the surface of the plasma membrane ('angle from PM') was measured using the angle tool in ImageJ by selecting three points: a point on the ribbon about 150 nm from the synapse, a point at the juncture of the ribbon and the synaptic density on the plasma membrane, and a point located on the plasma membrane about 150 nm from the juncture with the ribbon, producing an acute angle.

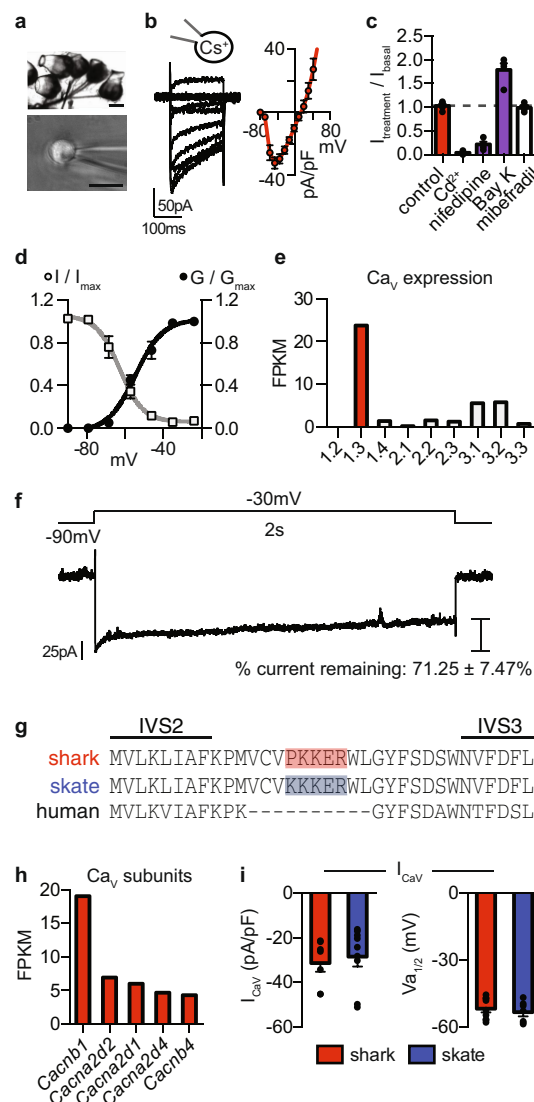
**Behavioural analysis.** In an isolated location and under normal lighting conditions, individual juvenile skates ( $n = 6$ ) and sharks ( $n = 5$ ) of both sexes were allowed to freely move and habituate for 20 min in an ambient temperature, seawater-filled cylindrical acrylic tank (diameter 28 cm). A sinusoidal electrical stimulus (100  $\mu$ A over 5 mm), generated by threading positive and negative ends of tin-plated copper wire (300 VH, 22 gauge, NTE Electronics, Inc.) into seawater-filled Tygon tubing powered by a function generator (Tone Generator Pro, Performance Audio) was randomly positioned and obscured by sand substrate in one of four circles (diameter 5.5 cm), all equally spaced from the centre of the tank. After an initial recording of baseline ventilation frequency, individual fish were stimulated at 2, 5, 10, 25, 50, 100, 150 and 200 Hz for 5 min, in randomized order. Each frequency was tested 10 times. All skates were exposed to a plume of Mysis shrimp odorant, and sharks were presented with squid odorant, to measure responses to natural food stimuli. To prevent habituation to the stimuli, an interval of 20 min without electrical stimuli was used between each trial. A digital video camera (Panasonic HC-V770) was positioned above the tank and used to measure ventilatory responses, as characterized by cyclical movement of spiracles or gills. Measurements were randomized and made blind to stimulation conditions.

**Statistical analysis.** Data were analysed with Clampfit (Axon Instruments), Patchmaster (HEKA) or Prism (Graphpad). Data are represented as mean  $\pm$  s.e.m. and *n* represents independent experiments for the number of cells in electrophysiology, quantified structures from histological analysis, or behavioural trials. Data were considered significant if  $P < 0.05$  using paired or unpaired two-tailed Student's *t*-tests or one- or two-way ANOVAs. All significance tests were justified considering the experimental design and we assumed normal distribution and variance, as is common for similar experiments. Sample sizes were chosen on the basis of the number of independent experiments required for statistical significance and technical feasibility.

**Reporting summary.** Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

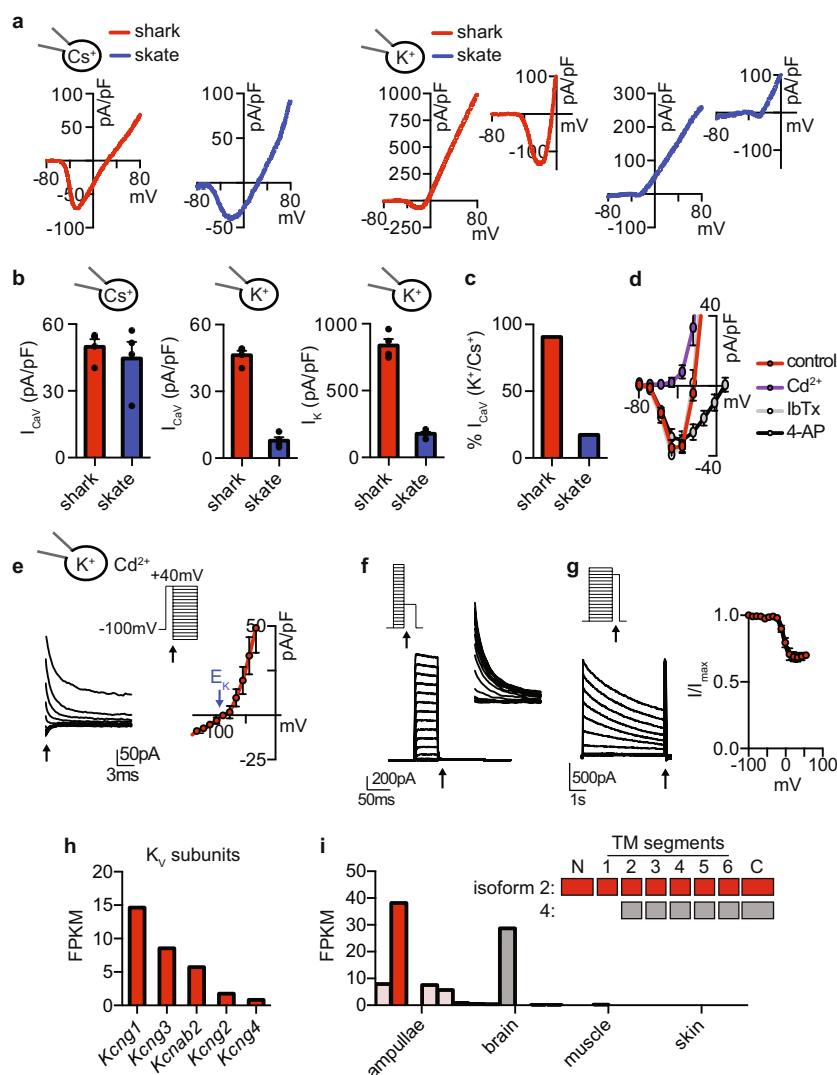
**Data availability.** Deep sequencing data are archived in the Gene Expression Omnibus under accession number GSE103977. The GenBank accession numbers for the  $\alpha$  subunit are: Cav1.3 MF959522, Kv1.3 MF959523 and BK MF959524. Other data are available from the corresponding author upon reasonable request.

27. Gillis, J. A., Dahn, R. D. & Shubin, N. H. Chondrogenesis and homology of the visceral skeleton in the little skate, *Leucoraja erinacea* (Chondrichthyes: Batoidea). *J. Morphol.* **270**, 628–643 (2009).
28. Sigg, D., Stefani, E. & Bezanilla, F. Gating current noise produced by elementary transitions in Shaker potassium channels. *Science* **264**, 578–582 (1994).
29. Yang, Y., Yan, Y. & Sigworth, F. J. How does the W434F mutation block current in Shaker potassium channels? *J. Gen. Physiol.* **109**, 779–789 (1997).
30. Ishii, T., Omura, M. & Mombaerts, P. Protocols for two- and three-color fluorescent RNA in situ hybridization of the main and accessory olfactory epithelia in mouse. *J. Neurocytol.* **33**, 657–669 (2004).
31. Kantardzhieva, A., Liberman, M. C. & Sewell, W. F. Quantitative analysis of ribbons, vesicles, and cisterns at the cat inner hair cell synapse: correlations with spontaneous rate. *J. Comp. Neurol.* **521**, 3260–3271 (2013).



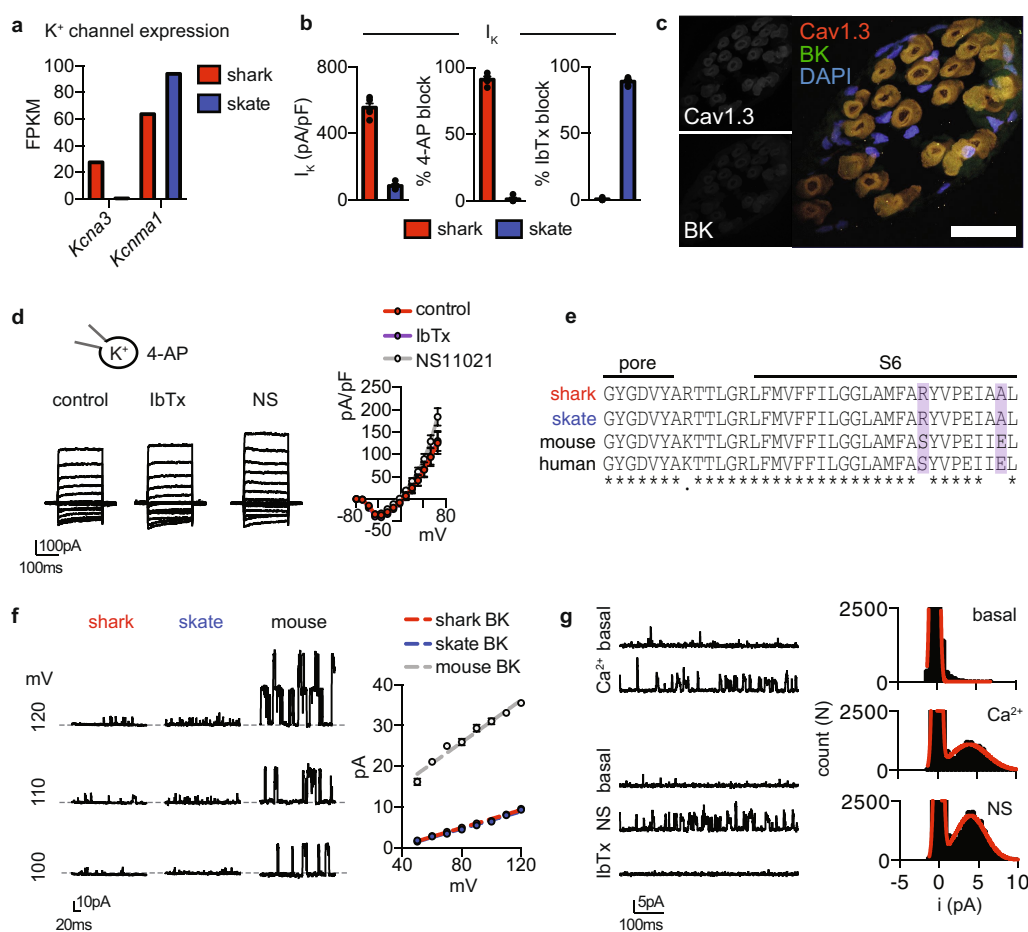
**Extended Data Fig. 1 | Properties of shark  $I_{\text{CaV}}$ .** **a**, Top, isolated shark ampullary organs with attached canals and nerve fibres (scale bar, 100  $\mu\text{m}$ ); bottom, a representative electrosensory cell patch-clamp experiment (scale bar, 10  $\mu\text{m}$ ). **b**, Left,  $I_{\text{CaV}}$  currents elicited by increasing voltage pulses from  $-100$  mV; right, average current-voltage ( $I$ - $V$ ) relationship ( $n = 7$ ). **c**,  $I_{\text{CaV}}$  exhibited an L-type  $\text{Ca}_v$  pharmacological profile: Peak currents were regulated by Bay K (agonist),  $\text{Cd}^{2+}$  (blocker) and nifedipine (antagonist), but not by mibefradil (T-type antagonist).  $n = 4$ .  $P < 0.0001$  for control versus all treatments except mibefradil, one-way ANOVA with post hoc Bonferroni test. **d**,  $I_{\text{CaV}}$  conductance-voltage ( $G$ - $V$ ) relationship (black) with half-maximal activation voltage ( $V_{1/2}$ ) of  $-54.6 \pm 1.2$  mV. Inactivation-voltage relationship (grey) with half-inactivation potential ( $V_{h1/2}$ ) of  $-62.9 \pm 1.4$  mV. A large window current was observed between

$-70$  mV and  $-40$  mV.  $G$ - $V$  relationships were established from current measurements during voltage pulses delivered in 10-mV increments from  $-100$  mV. Inactivation was measured at a  $-20$  mV test pulse after a series of voltage prepulses.  $n = 7$ . **e**,  $\text{Ca}_v$   $\alpha$ -subunit mRNA expression in shark ampullary organs. Bars represent FPKM. **f**,  $I_{\text{CaV}}$  elicited by a 2-s depolarizing voltage step to  $-30$  mV exhibits little inactivation. Representative of  $n = 5$ . **g**,  $\text{Ca}_v1.3$  alignment revealed that the IVS2-S3 motif that confers low voltage threshold in skate  $\text{Ca}_v1.3$  is conserved in the shark orthologue. **h**, Expression of  $\text{Ca}_v$  auxiliary subunits in shark ampullary organs. Bars represent FPKM. **i**, Average  $I_{\text{CaV}}$  current density and voltage-activation threshold was similar in shark and skate electrosensory cells.  $n = 6$ . All data are represented as mean  $\pm$  s.e.m.,  $n$  denotes cells.



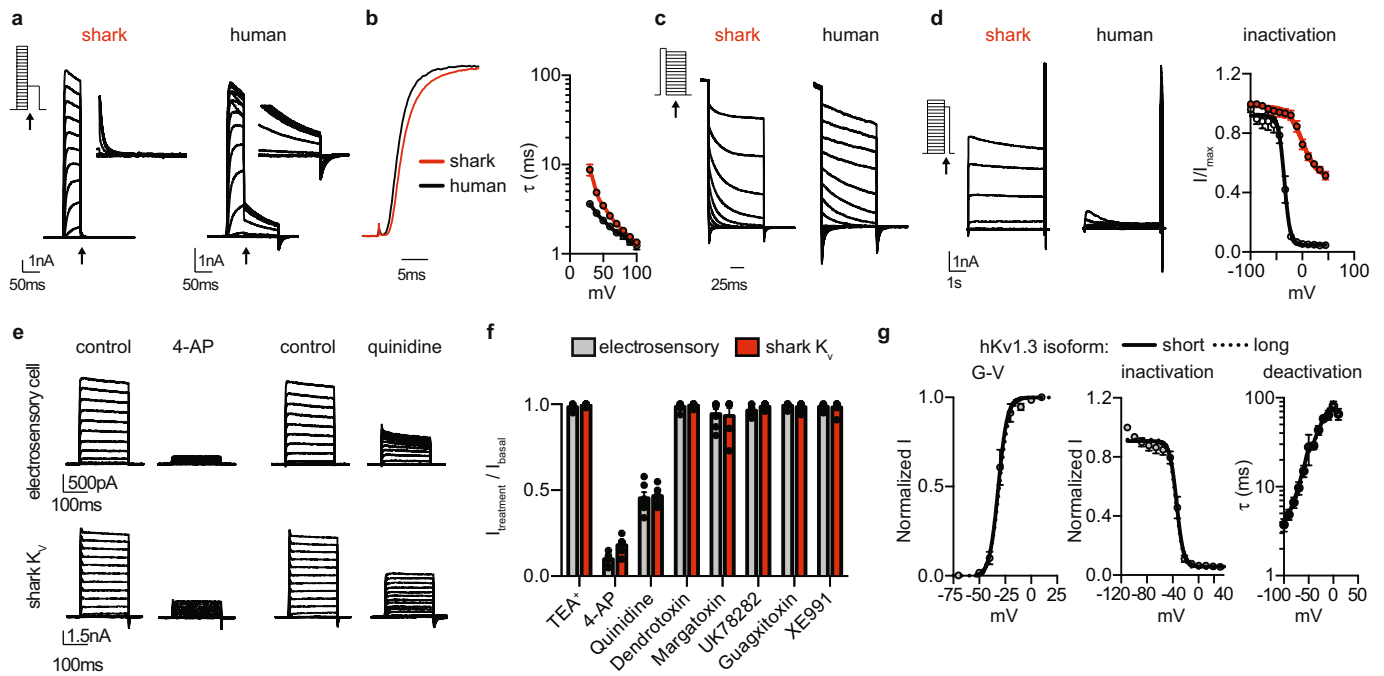
**Extended Data Fig. 2 | Properties of shark  $I_{\text{KV}}$ .** **a**, Currents elicited by 500-ms voltage ramps in shark (red) or skate (blue) electrosensory cells in the presence of intracellular  $\text{Cs}^+$  (left) or  $\text{K}^+$  (right). The insets on the right show the inward  $I_{\text{CaV}}$  in the presence of intracellular  $\text{K}^+$ . **b**, Average  $I_{\text{CaV}}$  current density elicited by voltage ramps was similar for both shark and skate cells in the presence of intracellular  $\text{Cs}^+$ , but larger in shark cells in the presence of intracellular  $\text{K}^+$ . Outward  $\text{K}^+$  current density was significantly larger in shark cells.  $n = 5$ ,  $P < 0.0001$  for shark versus skate  $I_{\text{CaV}}$  or  $I_{\text{K}}$  with intracellular  $\text{K}^+$ , two-tailed Student's  $t$ -test. **c**, The percentage of  $I_{\text{CaV}}$  remaining in the presence of  $\text{K}^+$  compared with  $\text{Cs}^+$  is markedly greater in shark cells compared with those from skate.  $n = 5$ ,  $P < 0.0001$  for inward control currents versus  $\text{Cd}^{2+}$ , two-way ANOVA with post hoc Tukey test. Peak inward currents were not affected by IbTx or 4-AP. **d**, Inward currents elicited by increasing voltage pulses from  $-100$  mV were not affected by IbTx or 4-AP, but were blocked by  $\text{Cd}^{2+}$ .  $n = 5$ ,  $P < 0.0001$  for inward control currents versus  $\text{Cd}^{2+}$ , two-way ANOVA with post hoc Tukey test. Peak inward currents were not affected by IbTx or 4-AP. **e**, Reversal potential for  $I_{\text{KV}}$  in shark electrosensory cells is near the reversal potential for selective  $\text{K}^+$  permeation ( $E_{\text{K}}$ , blue arrow on the

$I$ - $V$  plot).  $n = 5$ . Arrows indicate when currents were measured at the indicated voltages after an activating prepulse of 40 mV (also shown in expanded view). Extracellular  $\text{Cd}^{2+}$  was included to block  $I_{\text{CaV}}$  for biophysical studies of  $I_{\text{KV}}$ . **f**,  $I_{\text{KV}}$  currents elicited by a voltage protocol to obtain the  $G$ - $V$  relationship. Arrows indicate when tail currents were measured at  $-30$  mV after voltage steps that increased in 10-mV increments from  $-100$  mV (expanded view within inset). Representative of  $n = 11$ . **g**, Voltage-dependent inactivation properties of  $I_{\text{KV}}$ . The arrow indicates when inactivation was measured during 40-mV pulses after a series of prepulses that increased in 10-mV increments from  $-100$  mV.  $V_{h1/2}$  was  $-5.5 \pm 1.7$  mV and inactivation was incomplete.  $n = 6$ . **h**, mRNA expression of the  $\text{Kv}$  channel auxiliary subunits in shark ampullary organs. Bars represent FPKM. **i**, mRNA expression of *kcna3* isoforms. The major isoform studied is indicated in red. Other low-expression ampullary isoforms were similar. The only isoform that was appreciably expressed outside of ampullae was truncated and found in the brain (grey). All data are represented as mean  $\pm$  s.e.m.,  $n$  denotes cells.



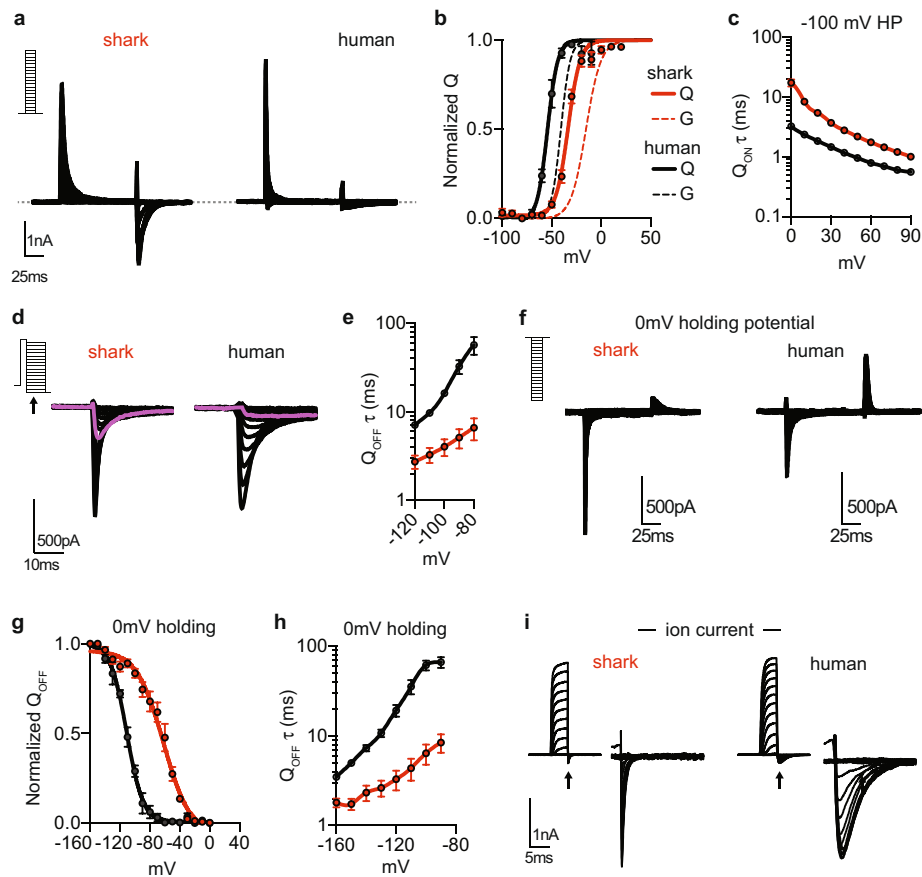
**Extended Data Fig. 3 | Properties of shark BK.** **a**, mRNA expression of major K<sup>+</sup>-channel  $\alpha$ -subunits (*Kcna3*, *Kv1.3*) and (*Kcnma1*, BK) in shark and skate electrosensory cells. **b**, Average K<sup>+</sup> current density, 4-AP-sensitive current ( $I_{KV}$ ), and IbTx-sensitive current ( $I_{BK}$ ) in shark and skate electrosensory cells.  $n = 5$ ,  $P < 0.0001$  for shark versus skate cells for all comparisons, two-tailed Student's  $t$ -test. **c**, Co-localization of Cav1.3 (red) and BK (green) transcripts within shark ampullary organs. Nuclei were stained with DAPI (blue). Scale bar, 100  $\mu$ m. Representative of  $n = 4$ . **d**, In the presence of 4-AP, a relatively small outward current remained that was insensitive to IbTx and was slightly increased by NS11021 at very positive voltages.  $n = 5$ ,  $P < 0.05$  for control versus NS11021 at 70 or 80 mV, two-way ANOVA with post hoc Tukey test. **e**, BK alignment revealed that residues found to reduce conductance in skate BK are conserved in

the shark orthologue. **f**, Heterologously expressed shark and skate BK had relatively small single-channel conductance compared with mouse BK. Shark BK =  $109 \pm 4$  pS; skate BK =  $105 \pm 5$  pS; mouse BK =  $259 \pm 12$  pS.  $n = 5$ ,  $P < 0.0001$  for mouse versus shark or skate BK, two-way ANOVA with post hoc Tukey test. **g**, 1  $\mu$ M Ca<sup>2+</sup> increased open probability in shark BK expressed in excised inside-out patches, and 10  $\mu$ M NS11021 increased open probability of channels in excised outside-out patches. NS11021 modulation was blocked by 100 nM IbTx. Holding voltage was 80 mV. NP<sub>o</sub>: basal,  $0.0061 \pm 0.0014$ ; Ca<sup>2+</sup>,  $0.11 \pm 0.011$ , NS11021,  $0.24 \pm 0.026$ .  $n = 5$ ,  $P < 0.0001$  for basal versus Ca<sup>2+</sup> or NS11021, two-tailed Student's  $t$ -test. All data are represented as mean  $\pm$  s.e.m,  $n$  denotes cells or tissue sections.



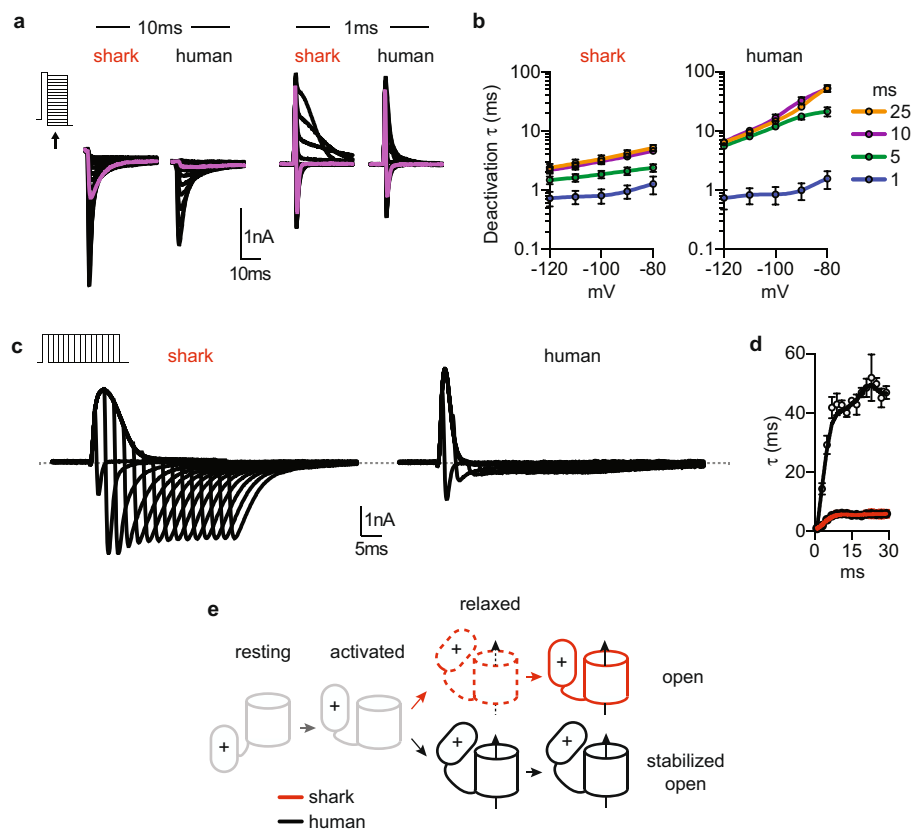
**Extended Data Fig. 4 | Properties of shark  $K_V$ .** **a**, Voltage-activated currents recorded in HEK293 cells expressing shark (red) or human (black)  $K_V1.3$ . Arrows indicate when currents were measured at  $-30$  mV after voltage pulses that increased in  $10$ -mV increments from  $-100$  mV. The inset shows expansions of measured currents following arrows. **b**, Left, normalized currents elicited by a  $40$ -mV voltage pulse demonstrated that expressed shark  $K_V$  channels open more slowly compared with human orthologues. Right, average activation kinetics were slower for shark  $K_V$  compared with human  $K_V$  in response to voltages from  $20$  mV to  $50$  mV.  $n = 6$ ,  $P < 0.0001$  for contribution of orthologue identity to series variance, two-way ANOVA with post hoc Bonferroni test. **c**, Deactivation kinetics of normalized currents from shark and human  $K_V$  channels at various repolarizing voltages after an activating prepulse of  $40$  mV. The arrow indicates when current properties were measured during the voltage protocol. **d**, Inactivation properties (left) and average inactivation-voltage relationships (right) of shark (red,  $V_{h1/2} = 0.1 \pm 2.8$  mV) and human (black,  $V_{h1/2} = -30.6 \pm 0.9$  mV)

$K_V1.3$  channels. The arrow indicates when inactivation was measured during  $40$ -mV pulses after a series of prepulses that increased in  $10$ -mV increments from  $-100$  mV.  $n = 9$ ,  $P < 0.0001$  for  $V_{h1/2}$ , two-tailed Student's  $t$ -test. **e**,  $I_{KV}$  was reduced by  $4$ -AP or quinidine in native shark electrosensory cells or heterologously expressed shark  $K_V1.3$ . Currents were elicited by increasing voltage pulses from  $-90$  mV (native) or  $-100$  mV (heterologous). **f**, Pharmacological profile of shark electrosensory cell  $I_{KV}$  and heterologously expressed shark  $K_V1.3$ . Currents measured at peak amplitude were reduced by  $4$ -AP or quinidine, but not by other treatments. Pharmacological modulation of native  $I_{KV}$  and shark  $K_V1.3$  was similar. **g**, The short human isoform of  $K_V1.3$  (short N-terminal truncation) was used to study gating currents because of enhanced expression, but channel properties are identical<sup>15</sup>. Similarly, we found that activation threshold and  $G$ - $V$  relationship, voltage-dependent inactivation, and deactivation were similar between long and short isoforms.  $n = 6$ . All data are represented as mean  $\pm$  s.e.m,  $n$  denotes cells.



**Extended Data Fig. 5 | Shark  $K_v$  gating currents.** **a**, Gating currents recorded in HEK293 cells expressing non-conductive shark (red) or human (black)  $K_v1.3$  elicited by increasing voltage pulses in 10-mV increments from a  $-100$ -mV holding potential. **b**, Average charge ( $Q$ )– $V$  relationships. Shark  $K_v1.3$   $V_{a1/2}$  was  $-33.4 \pm 0.5$  mV,  $K_a = 6.1 \pm 0.5$  mV; and human  $K_v1.3$   $V_{a1/2}$  was  $-54.25 \pm 0.7$  mV,  $K_a = 5.2 \pm 0.6$  mV.  $n = 9$ ,  $P < 0.0001$  for  $V_{a1/2}$  two-tailed Student's  $t$ -test. Dotted lines indicate associated  $G$ – $V$  relationships. **c**,  $Q_{ON}$  kinetics after voltage sensor activation from a holding potential of  $-100$  mV were significantly slower in shark (red) compared with human (black).  $n = 7$ ,  $P < 0.0001$  for contribution of orthologue identity to series variance, two-way ANOVA with post hoc Bonferroni test. **d**, Representative OFF gating current ( $Q_{OFF}$ ) kinetics during repolarization in 10-mV increments after a 40-mV prepulse. The arrow indicates when deactivation rates were measured and purple traces show deactivation at  $-50$  mV. **e**,  $Q_{OFF}$  kinetics were significantly faster in shark (red) compared with human (black).  $n = 11$ ,  $P < 0.0001$  for contribution of orthologue identity to series variance, two-way ANOVA with post hoc Bonferroni test. **f**, Gating currents from shark (red) or human (black)  $K_v1.3$  after

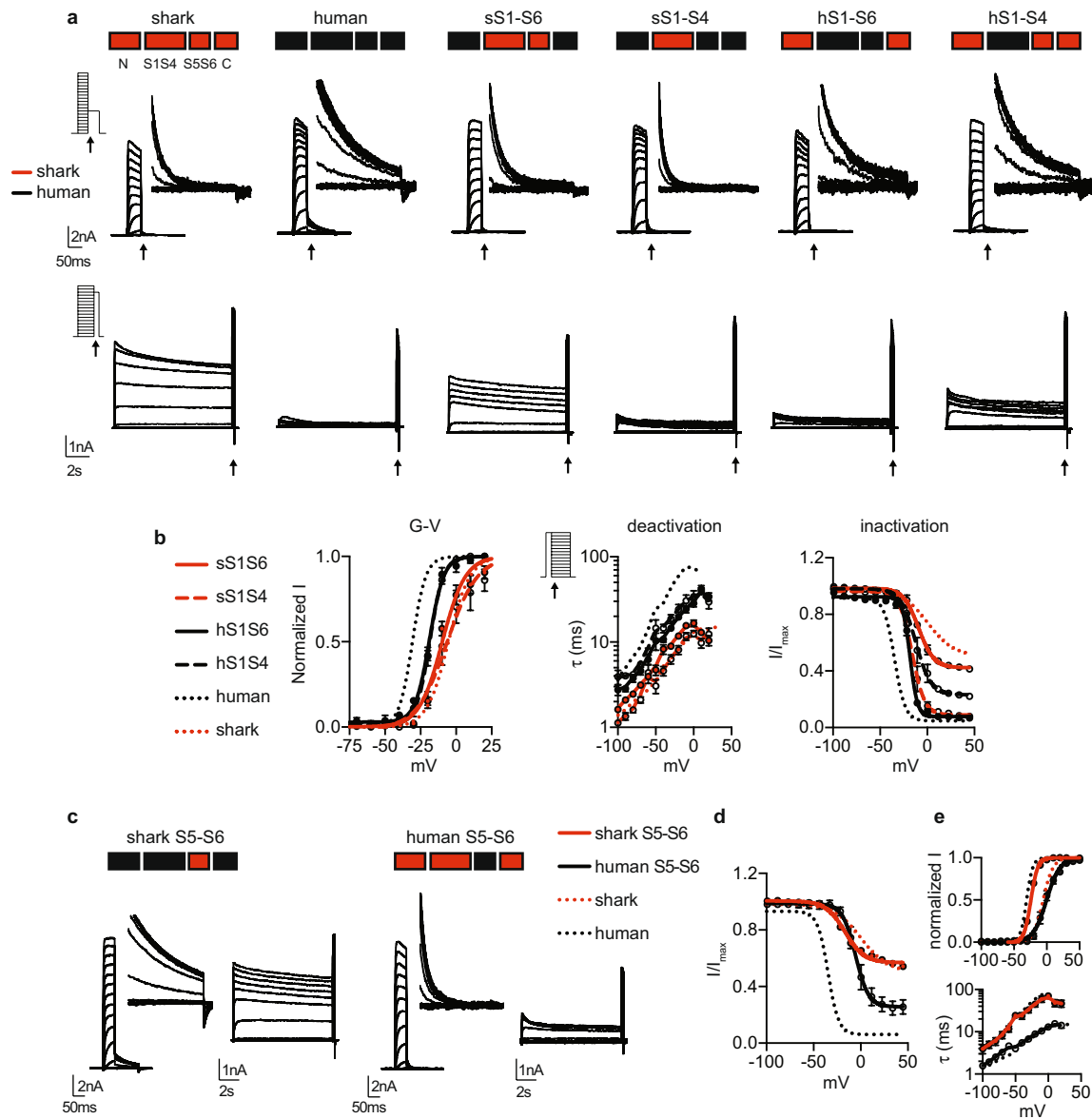
decreasing voltage pulses in increments of 10 mV from a holding potential of 0 mV. Scale bars, 500 pA, 25 ms. **g**, Average charge ( $Q$ )– $V$  relationships of downward voltage sensor movement ( $Q_{OFF}$ ) in response to decreasing voltage pulses from a holding potential of 0 mV. Shark  $K_v1.3$   $V_{a1/2}$  was  $-61.6 \pm 1.9$  mV and human  $K_v1.3$   $V_{a1/2}$  was  $-110.9 \pm 1.01$  mV.  $n = 7$ ,  $P < 0.0001$  for contribution of orthologue identity to series variance, two-way ANOVA with post hoc Bonferroni test. **h**,  $Q_{OFF}$  kinetics of voltage sensor deactivation from a holding potential of 0 mV were significantly faster in shark (red) compared with human (black).  $n = 7$ ,  $P < 0.0001$  for contribution of orthologue identity to series variance, two-way ANOVA with post hoc Bonferroni test. All data are represented as mean  $\pm$  s.e.m. **i**, Ion tail currents (indicating channel closure) deactivated faster in shark (red) compared with human (black)  $K_v1.3$ . Tail currents were measured at  $-100$  mV after a series of activating voltage pulses that increased in 10-mV increments. Inset, arrows indicate when tail currents were measured after activating voltage pulses. Representative of  $n = 10$ .  $\tau$  for deactivation from 60-mV pulse: shark =  $1.5 \pm 0.1$  ms; human =  $5.5 \pm 0.4$  ms.  $P < 0.0001$ , two-tailed Student's  $t$ -test. All data are represented as mean  $\pm$  s.e.m.,  $n$  denotes cells.



#### Extended Data Fig. 6 | Shark $K_V$ voltage sensor domain relaxation.

**a**,  $Q_{OFF}$  kinetics after either 10-ms or 1-ms activating prepulses of 40 mV. Purple traces indicate deactivation at -50 mV. Arrow indicates when current properties were measured during the voltage protocol. **b**, Average  $Q_{OFF}$  kinetics were faster in shark compared with human during deactivation after 40-mV prepulses of 25, 10, or 5 ms duration, but rates were similar after 1-ms activating prepulses. Kinetics were measured at voltages that decreased in 10-mV increments from 40 mV.  $n = 9$ ,  $P < 0.0001$  for contribution of orthologue identity to series variance at 25, 10, or 5 ms, but no significant difference was observed at 1 ms, two-way ANOVA with post hoc Bonferroni test. **c**, Shark  $K_V1.3$   $Q_{OFF}$  kinetics were relatively unaffected by the duration of the activating voltage pulse, whereas human  $K_V1.3$  entered a proposed 'relaxed' state that resulted

in the slowing of  $Q_{OFF}$  with increasing pulse length. Deactivation was measured at -100 mV after a series of 40-mV voltage pulses of varying duration from 0.5 ms to 30 ms. **d**, Average  $Q_{OFF}$  kinetics in response to indicated voltage pulse lengths.  $n = 6$ ,  $P < 0.0001$  for comparison of orthologue  $Q_{OFF}$  kinetics after a 30-ms voltage pulse. **e**, Hypothetical model of shark and human  $K_V1.3$ . Compared with its human orthologue, shark  $K_V1.3$  exhibits reduced voltage sensor domain relaxation, which stabilizes pore opening in human  $K_V1.3$ . Reduced voltage sensor relaxation is indicated by dotted lines to suggest that this state (or states) may occur to a lesser extent in the shark orthologue. Thus, compared with human  $K_V1.3$ , the shark orthologue requires relatively less repolarizing voltage to more quickly return to a resting/closed state. All data are represented as mean  $\pm$  s.e.m,  $n$  denotes cells.



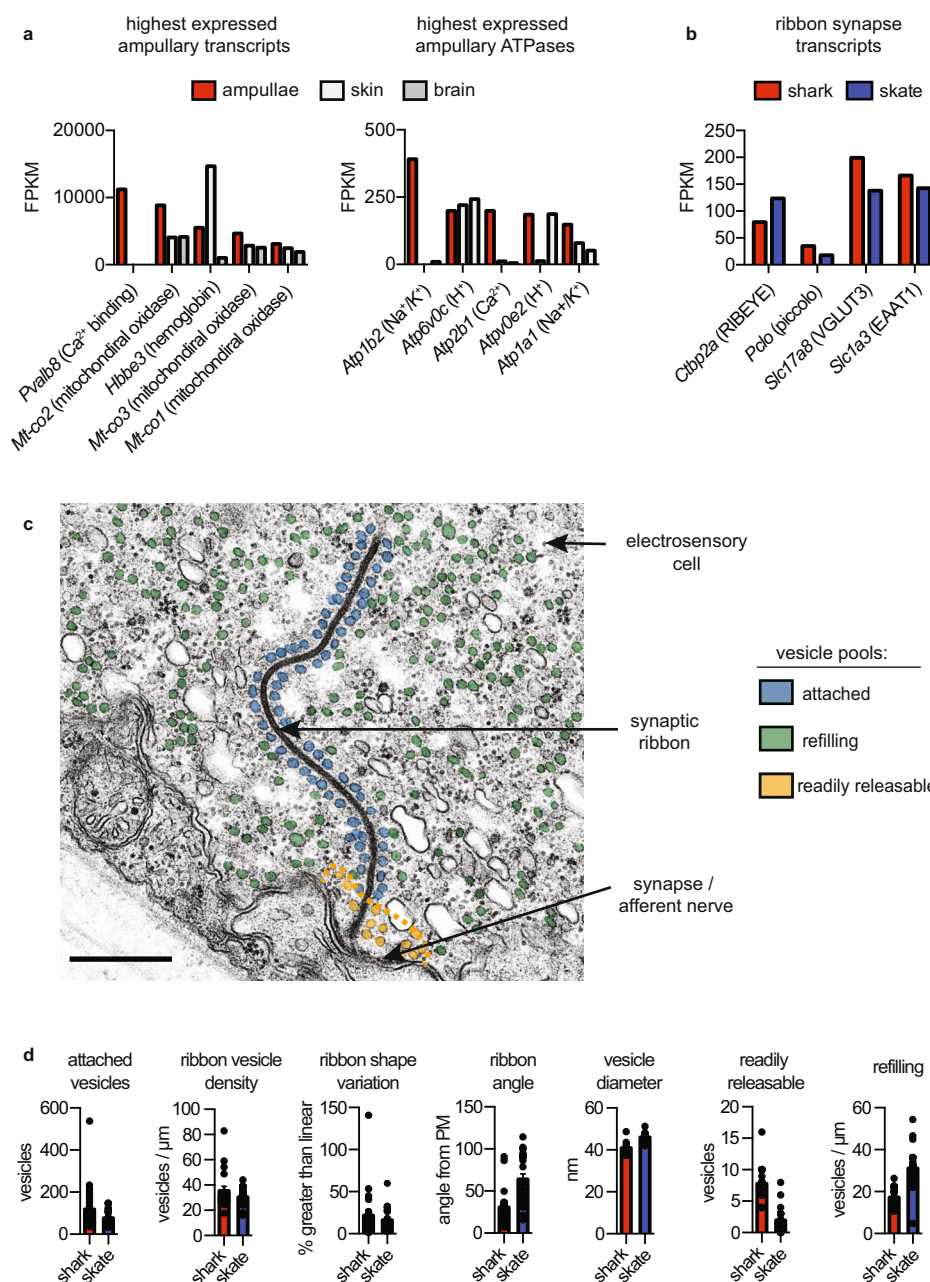
### Extended Data Fig. 7 | Shark-human $K_v$ chimaeric analyses.

**a**, Chimaeric shark-human  $K_v1.3$  channels reveal that shark  $K_v$  S1-S6 confers differences in activation voltage threshold, deactivation kinetics and inactivation. Top, the chimaera constructs analysed (shark in red, human in black). Middle, the arrow indicates when voltage-activated currents were measured at  $-30$  mV after a series of voltage pulses that increased in 10-mV increments from  $-100$  mV. Bottom, the arrow indicates when inactivation was measured during 40-mV pulses after a series of prepulses that increased in 10-mV increments from  $-100$  mV.

**b**, Compared with wild-type human  $K_v$ , average  $G-V$  relationships for wild-type shark, shark (s)S1-S6, and sS1-S4 channels were similarly shifted to positive voltages with more gradual slopes and deactivation kinetics were faster.  $V_{a1/2}$  (mV) for wild-type human =  $-30.7 \pm 0.5$  mV, slope factor ( $K_a$ ) =  $4.7 \pm 0.5$  mV; wild-type shark  $V_{a1/2}$  =  $-5.4 \pm 0.5$  mV,  $K_a$  =  $7.6 \pm 0.4$  mV; sS1-S6  $V_{a1/2}$  =  $-9.7 \pm 0.7$  mV,  $K_a$  =  $8.1 \pm 0.6$  mV; sS1-S4  $V_{a1/2}$  =  $-6.5 \pm 0.7$  mV,  $K_a$  =  $10.7 \pm 1.2$  mV. Average deactivation kinetics of wild-type shark, sS1-S6, and sS1-S4  $K_v$  channels were faster than those of wild-type human channels. Substitution of human (h)S1-S6 or hS1-S4 into the shark  $K_v$  channel also partially shifted the activation

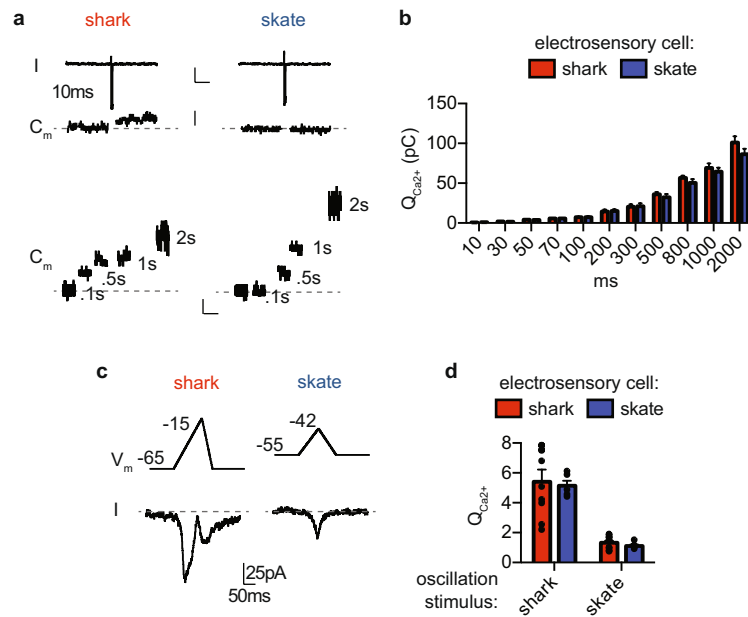
threshold and deactivation kinetics. Channels containing hS5-S6 exhibited the strongest voltage-dependent inactivation—nearly as efficient as wild-type human  $K_v$ —whereas channels containing sS5-S6 displayed weaker inactivation. sS1-S4 had a smaller effect on inactivation.  $V_{h1/2}$  for wild-type human =  $-34.6 \pm 0.9$  mV, wild-type shark =  $0.1 \pm 2.8$  mV, hS1-S6 =  $-18.3 \pm 0.4$  mV, hS1-S4 =  $-12.3 \pm 1.2$  mV, sS1-S6 =  $-9.2 \pm 0.8$  mV, sS1-S4 =  $-16.0 \pm 0.9$  mV.  $n = 9$  for each of wild-type shark, wild-type human, hS1-S6, hS1-S4 and sS1-S6, and 7 for sS1-S4.

**c**, Currents elicited from the indicated S5-S6 chimaeric channels in response to voltage protocols to access voltage-dependence for activation and inactivation. **d**, sS5-S6 reduces voltage-dependent inactivation and hS5-S6 partially confers strong voltage-dependent inactivation in shark channels.  $n = 6$ .  $P < 0.0001$  for hS5-S6 versus sS5-S6 or wild-type shark, sS5-S6 versus hS5-S6 or wild-type human, two-way ANOVA with post hoc Tukey test. **e**, Top, S5-S6 substitution did not greatly affect voltage-dependent activation.  $n = 6$  for hS5-S6, 7 for sS5-S6. Bottom, S5-S6 substitution did not affect deactivation kinetics.  $n = 7$  for hS5-S6, 9 for sS5-S6. All data are represented as mean  $\pm$  s.e.m.,  $n$  denotes cells.



**Extended Data Fig. 8 | Electrosensory cell ribbon synapse characteristics.** **a**, Left, five highest expressed transcripts in shark ampullae. The  $\text{Ca}^{2+}$ -binding protein parvalbumin 8 is the most highly expressed and is enriched in ampullae compared with other examined tissues. Right, five highest expressed ATPase transcripts in shark ampullae. A plasma membrane  $\text{Ca}^{2+}$  ATPase is highly expressed and enriched in ampullae. Bars represent FPKM. **b**, Expression of transcripts associated with ribbon synapses in shark and skate ampullae. Expression of vGluT3 and EAAT1 suggests that the synapse could be glutamatergic. **c**, Transmission electron micrograph of skate ribbon synapse with arrows indicating electrosensory cell, synaptic ribbon and afferent nerve terminal. Distinct vesicular pools are coloured: blue, attached to ribbon; green, refilling; yellow, readily releasable. An orange dotted line indicates the 150-nm region in which the readily releasable pool was quantified. Scale

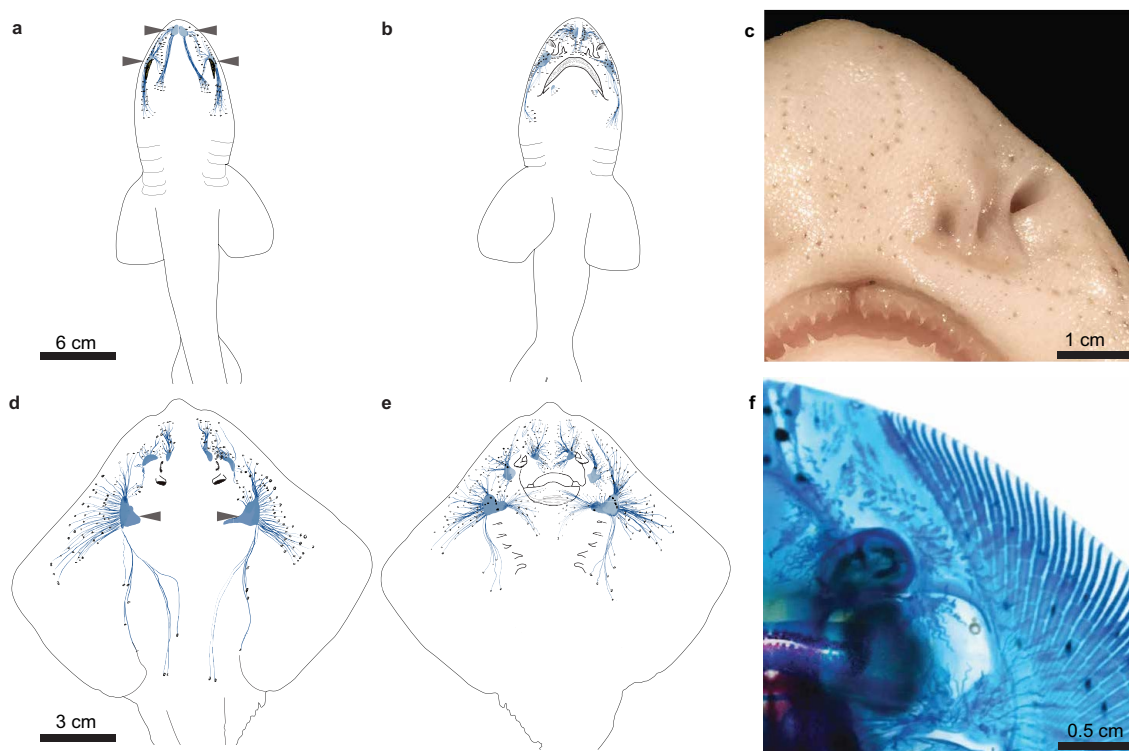
bar, 500 nm. **d**, Quantification of attached vesicles, ribbon vesicle density, ribbon shape variation and vesicle diameter was similar between shark and skate electrosensory cells. The readily releasable pool, quantified by number of vesicles 150 nm from the synapse, was significantly larger in shark versus skate electrosensory cells.  $n = 18$ ,  $P < 0.0001$ , two-tailed Student's  $t$ -test. The refilling pool density, quantified as detached cytosolic vesicles, was significantly larger in skate electrosensory cells.  $n = 20$  shark and 21 skate,  $P < 0.0001$ , two-tailed Mann-Whitney test. Shark ribbons were more parallel to the plasma membrane in comparison to skate ribbons that were often more perpendicular. For angle quantification  $n = 20$  shark and 21 skate ribbons,  $P < 0.0001$ , two-tailed Mann-Whitney test. All data are represented as mean  $\pm$  s.e.m,  $n$  denotes counted structures.



#### Extended Data Fig. 9 | Shark electrosensory cell vesicular release characteristics.

**a**, Top, currents and capacitance changes in response to a 10 ms,  $-20$  mV voltage pulse in shark and skate electrosensory cells. Scale bars, 50 pA, 200 ms. Bottom, representative capacitance changes in response to the indicated durations of a voltage stimulus of  $-20$  mV. Scale bars, 25 fF, 200 ms. **b**,  $-20$  mV voltage pulses of various durations induced similar integrated  $I_{CaV}$  ( $Q_{Ca^{2+}}$ ) in shark or skate electrosensory cells.  $n = 6$ . **c**,  $I_{CaV}$  elicited by simulated voltage spikes in shark electrosensory cells and smaller voltage oscillations in skate cells.  $K^+$  currents were blocked by

intracellular  $Cs^+$ , extracellular 4-AP and IbTx. **d**, Voltage-clamp protocols developed to simulate shark electrosensory cell spiking induced the same amount of  $Q_{Ca^{2+}}$  in shark or skate cells. Similarly, voltage protocols that simulated smaller skate electrosensory cell voltage oscillations induced the same amount of  $Q_{Ca^{2+}}$  in shark or skate cells.  $Q_{Ca^{2+}}$  elicited by simulated voltage spikes was larger than  $Q_{Ca^{2+}}$  elicited by simulated oscillations.  $n = 10$  for shark and 5 for skate. All data are represented as mean  $\pm$  s.e.m,  $n$  denotes cells.



**Extended Data Fig. 10 | Schematic representation of ampullae of Lorenzini distribution in two elasmobranch species.** **a**, The dorsal surface of the chain catshark (*S. retifer*), with black dots corresponding to individual ampullary pores and blue lines representing canal structures. The buccal and supraorbital clusters from which electrosensory cells were obtained are indicated with arrowheads. This schematic was prepared from photographs of four individual fish. **b**, The ventral surface of

the catshark. **c**, Photograph of ampullary pores, visible on the ventral rostrum of an adult catshark. **d**, The dorsal surface of the little skate (*L. erinacea*). The hyoid capsules from which electrosensory cells were obtained are indicated with arrowheads. This schematic was prepared from photographs of four individual fish. **e**, The ventral surface of the skate. **f**, Photograph of a cleared Alcian blue-stained skate, revealing the ampullary canals from the ventral surface of the skate.

# The coding of valence and identity in the mammalian taste system

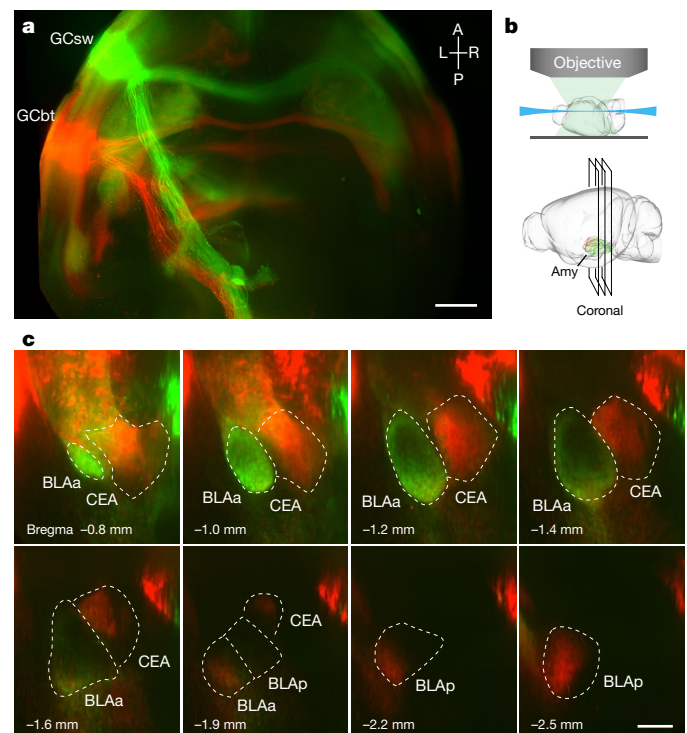
Li Wang<sup>1,2,3</sup>, Sarah Gillis-Smith<sup>1,2,3</sup>, Yueqing Peng<sup>1,2,3</sup>, Juen Zhang<sup>1,2,3</sup>, Xiaoke Chen<sup>1,2,6</sup>, C. Daniel Salzman<sup>3,4</sup>, Nicholas J. P. Ryba<sup>5</sup> & Charles S. Zuker<sup>1,2,3\*</sup>

**The ability of the taste system to identify a tastant (what it tastes like) enables animals to recognize and discriminate between the different basic taste qualities<sup>1,2</sup>. The valence of a tastant (whether it is appetitive or aversive) specifies its hedonic value and elicits the execution of selective behaviours. Here we examine how sweet and bitter are afforded valence versus identity in mice. We show that neurons in the sweet-responsive and bitter-responsive cortex project to topographically distinct areas of the amygdala, with strong segregation of neural projections conveying appetitive versus aversive taste signals. By manipulating selective taste inputs to the amygdala, we show that it is possible to impose positive or negative valence on a neutral water stimulus, and even to reverse the hedonic value of a sweet or bitter tastant. Remarkably, mice with silenced neurons in the amygdala no longer exhibit behaviour that reflects the valence associated with direct stimulation of the taste cortex, or with delivery of sweet and bitter chemicals. Nonetheless, these mice can still identify and discriminate between tastants, just as wild-type controls do. These results help to explain how the taste system generates stereotypic and predetermined attractive and aversive taste behaviours, and support the existence of distinct neural substrates for the discrimination of taste identity and the assignment of valence.**

The taste system is responsible for detecting and responding to the five basic taste qualities: sweet, sour, bitter, salty and umami<sup>1,2</sup>. Each of these five tastes is detected by specialized taste receptor cells on the tongue and palate epithelium, with different taste receptor cells dedicated to each of the taste modalities<sup>1,2</sup>. In rodents, taste information travels from taste receptor cells in the oral cavity to primary gustatory cortex (insular cortex) via four neural stations<sup>1,3</sup>: taste receptor cells to taste ganglia, then to the nucleus of the solitary tract, the parabrachial nucleus, the thalamus and to insular cortex. Intrinsic<sup>4,5</sup> and two-photon<sup>6</sup> imaging studies have shown that sweet and bitter taste are represented in the cortex in topographically separate cortical fields; by optogenetically activating these taste cortical fields in awake mice, it is possible to evoke prototypical taste behaviours in the total absence of taste stimuli<sup>7</sup>.

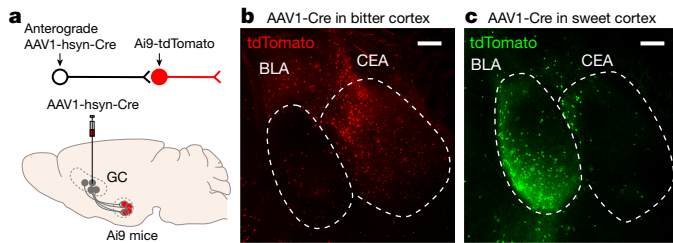
The two most important sensory features of a taste stimulus are its identity and its valence. We hypothesized that by examining the neural targets of the sweet and bitter cortical fields it may be possible to uncover the circuit logic for appetitive versus aversive tastes. To trace the projections of neurons in the sweet and bitter cortex, we labelled neurons in the sweet cortical field with enhanced green-fluorescent protein (eGFP), those in the bitter cortex with red-fluorescent protein (tdTomato), and then the whole brains were examined by clearing and rapid 3D imaging with light-sheet microscopy using clear, unobstructed brain imaging and computational analysis (CUBIC)<sup>8</sup>. Our results show that projections from the sweet and bitter cortical fields target multiple brain areas, including contralateral taste cortex, amygdala, entorhinal cortex, caudoputamen and thalamus (see Fig. 1). Notably, sweet and

bitter cortical projections exhibited strong segregation as separate lines while navigating to the amygdala (Fig. 1b, c and Extended Data Fig. 1), with neurons from the sweet cortical field terminating in the anterior basolateral amygdala (BLA), whereas neurons from the bitter cortical field predominantly projected to central amygdala (CEA), with some



**Fig. 1 | Projections from the sweet cortex and the bitter cortex terminate in distinct targets in the amygdala.** **a**, Maximum-intensity z stack of projections<sup>8</sup> from the sweet cortical field labelled with eGFP (GCsw, green) and the bitter cortical field labelled with tdTomato (GCbt, red). A, anterior; P, posterior; L, left; R, right. Scale bar, 1 mm. **b**, Schematic of whole-brain imaging with light-sheet fluorescence microscopy<sup>8</sup>, illustrating the coronal sections shown in **c**. Amy, amygdala. The brain diagrams were rendered by the scalable brain composer ([https://scalablebrainatlas.incf.org/services/sba-composer.php?template=ABA\\_v3](https://scalablebrainatlas.incf.org/services/sba-composer.php?template=ABA_v3)) based on Allen Mouse Brain Common Coordinate Framework version 3<sup>26,27</sup>. **c**, Segregation of sweet and bitter projections. Sweet cortical neurons project to the anterior BLA (BLAa, green), whereas bitter cortical neurons predominantly innervate the CEA (red) and a portion of the posterior BLA (BLAp, red; see also Extended Data Fig. 1). Scale bar, 0.5 mm. The boundaries of amygdala nuclei were based on the Allen Brain Institute atlas<sup>26</sup> (<http://brain-map.org/>). Similar results were observed in six animals.

<sup>1</sup>Howard Hughes Medical Institute, Columbia University, New York, NY, USA. <sup>2</sup>Department of Biochemistry and Molecular Biophysics, Columbia College of Physicians and Surgeons, Columbia University, New York, NY, USA. <sup>3</sup>Department of Neuroscience, Columbia College of Physicians and Surgeons, Columbia University, New York, NY, USA. <sup>4</sup>Department of Psychiatry and New York State Psychiatric Institute, Columbia University, New York, NY, USA. <sup>5</sup>National Institute of Dental and Craniofacial Research, National Institutes of Health, Bethesda, MD, USA. Present address: <sup>6</sup>Department of Biology, Stanford University, Stanford, CA, USA. \*e-mail: cz2195@columbia.edu



**Fig. 2 | Segregation of sweet and bitter targets in the amygdala.** **a**, Schematic illustrating anterograde transsynaptic labelling of neurons in the amygdala following AAV1-hsyn-Cre injection<sup>9</sup> in the sweet or bitter taste cortex of mice expressing a tdTomato reporter. **b**, **c**, Representative confocal images of tdTomato expression in amygdala following AAV1 injection in the bitter cortex (**b**) or sweet cortex (**c**; pseudocoloured green). Scale bars, 200  $\mu$ m. Similar results were obtained in three animals for each experiment. See also Extended Data Fig. 2.

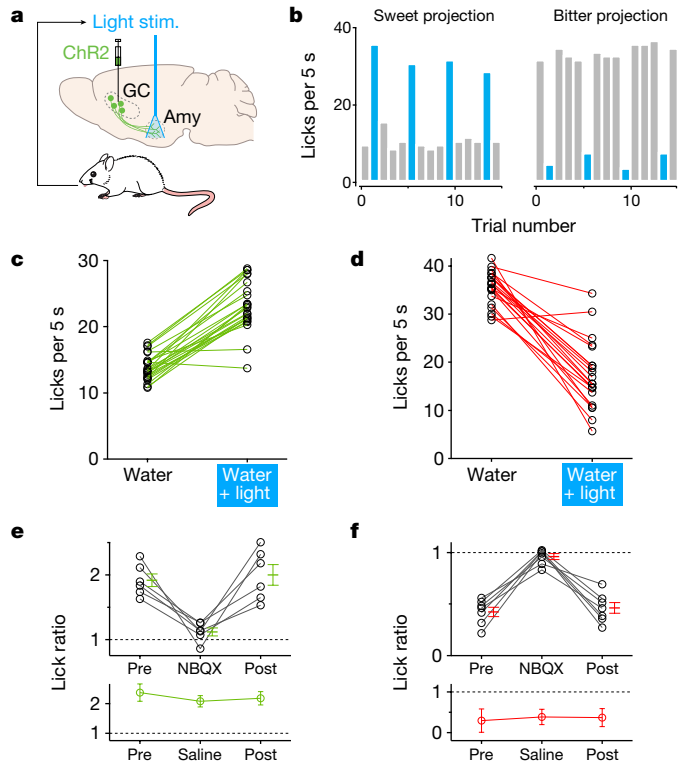
terminals in the posterior BLA. We extended these findings by performing anterograde labelling experiments using adeno-associated virus (AAV)-based transsynaptic transfer of Cre-recombinase<sup>9</sup> from sweet and bitter cortex to targets in the amygdala. These results substantiated that the BLA was the target of sweet cortex projections, and the CEA was the target of bitter cortex projections (Fig. 2; see Extended Data Fig. 2 for activity-dependent labelling).

The amygdala is a key brain structure involved in processing emotions, motivation and positive and negative stimuli<sup>10–19</sup>. Previous studies have shown that the BLA and CEA both contain distinct populations of neurons that are activated by negative or positive stimuli<sup>10,13–19</sup>. Our finding of such strong segregation of appetitive (sweet) versus aversive (bitter) projections to the amygdala immediately suggests an anatomical division for the generation of valence-specific behavioural responses to tastants.

If the amygdala imposes valence on tastants (that is, it represents the hedonic value of a tastant to drive valence-specific behaviours), then optogenetic activation of the terminals of sweet cortical neurons in the BLA should elicit attractive responses, whereas activation of bitter projections should evoke aversive behaviours. Therefore, we generated mice expressing channelrhodopsin-2 (ChR2)<sup>20</sup> in either the sweet or bitter cortical field, implanted optical fibres over the amygdala, and used a place-preference test to measure responses to photostimulation of the cortico-amygdalar projections. Our results showed that mice avoided the chamber linked to photostimulation of the bitter cortico-amygdalar projections (Extended Data Fig. 3), but exhibited a strong preference for the chamber associated with stimulation of the sweet projections.

Next, we reasoned that optogenetic activation of the terminals of sweet cortical neurons in the BLA would trigger appetitive taste behaviours, whereas stimulation of the projections from bitter cortical neurons in the CEA would instead impose a negative valence on the stimulus. Therefore, we assayed whether ChR2 activation of sweet-to-BLA projections while a mouse is drinking a neutral stimulus (for example, water) transforms it into a highly attractive one such as sugar, and conversely, whether activation of the projections from bitter cortex to CEA trigger strong laser-dependent suppression of licking, much like the introduction of a bitter chemical would do.

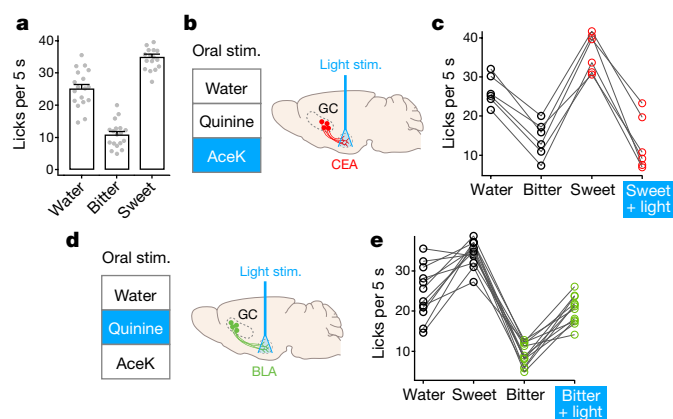
We used a behavioural paradigm in which ChR2-expressing mice were assayed for water drinking in a head-restrained setup<sup>7</sup>. In these experiments, the laser shutter was placed under lick-contact operation, and thus the mouse has control of its own stimulation during the light-on trials, and only self-stimulation would continue to trigger appetitive responses (light stimulation on its own does not trigger licking, or licking-like motor responses; see Methods<sup>7</sup>). By contrast, a mouse would immediately terminate licking if contact-licking elicited aversion. Indeed, optogenetic activation of the sweet cortex terminals in BLA evoked a marked increase in licking (self-stimulation; Fig. 3b, c), whereas activation of the bitter cortical projections to amygdala



**Fig. 3 | Activation of sweet and bitter cortical terminals in the amygdala drives appetitive and aversive behaviours.** **a**, Optogenetic stimulation strategy. Sweet neurons in the anterior gustatory cortex (GC) or bitter neurons in posterior gustatory cortex were transduced with AAV-ChR2. Stimulating optical fibres were placed above BLA or CEA. For coupling the photostimulation to drinking behaviour, laser pulses were triggered by licking. **b**, Representative histograms showing licking events in the presence (blue) or absence (grey) of photostimulation of sweet cortico-amygdalar projections (left) or bitter cortico-amygdalar projections (right). Note the marked enhancement or suppression of licking, respectively. **c**, **d**, Quantification of licking responses with and without light stimulation. **c**, Sweet cortical terminals in the BLA.  $n = 24$  mice, two-tailed paired  $t$ -test,  $P < 0.0001$ . **d**, Bitter cortical terminals in the CEA.  $n = 21$  mice, two-tailed paired  $t$ -test,  $P < 0.0001$ . See also Extended Data Fig. 4. **e**, **f**, Pharmacological silencing demonstrated that the light-dependent licking behaviours are due to activity in the amygdala; the panels show quantification of lick ratios before and after infusion of NBQX (top) or control saline (bottom) in the amygdala. **e**, Stimulation of sweet projections.  $n = 6$  mice. **f**, Stimulation of bitter projections.  $n = 7$  mice. Note that NBQX abolishes the light-dependent changes in licking responses. Values are mean  $\pm$  s.e.m. Repeated-measures one-way analysis of variance (ANOVA) followed by Bonferroni post hoc test (Supplementary Table 1).

strongly suppressed licking responses (Fig. 3b, d). To confirm that these light-triggered behaviours were not caused by back-propagation of action potentials from the stimulation in the amygdala (that is, back to the taste cortex and thus to other potential taste cortical targets), we repeated the experiment, however, this time we pharmacologically silenced synaptic activity locally in the amygdala by infusion of the AMPA ( $\alpha$ -amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid) receptor antagonist NBQX<sup>21</sup>. Our results demonstrate that silencing synaptic transmission in the amygdala abolished all light-evoked responses (Fig. 3e, f). As expected, responses fully recovered after wash-out of the drug (Fig. 3e, f). Taken together, these data demonstrate that activation of sweet or bitter cortico-amygdalar pathways is sufficient to impose a positive or a negative valence on a neutral taste cue.

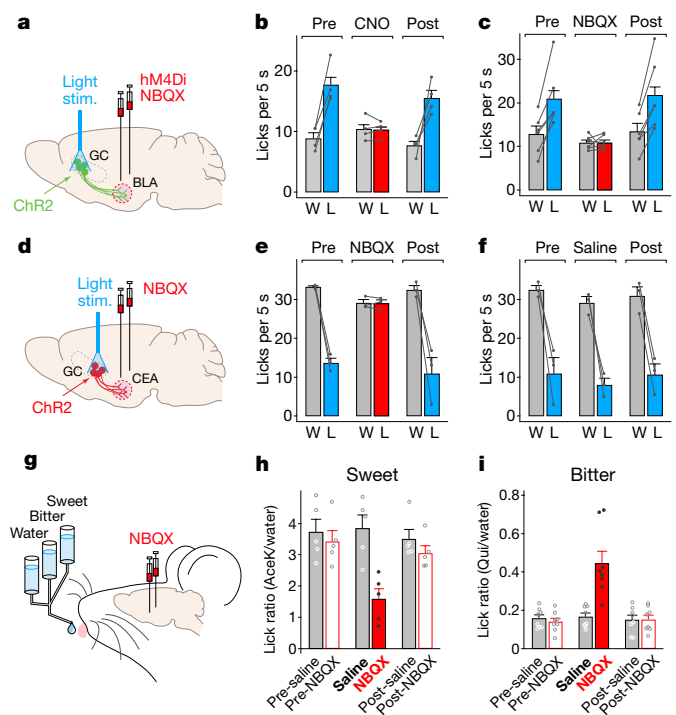
We hypothesized that strong activation of the bitter and sweet cortico-amygdalar projections might override the hedonic response elicited by sweet and bitter tastants. Therefore, we predicted that optogenetic activation of the bitter cortical terminals in the CEA may impose an



**Fig. 4 | Activation of cortico-amygdalar circuits overrides the hedonic valence of orally delivered tastants.** **a**, Licking responses (no photostimulation) to water, bitter (0.5 mM quinine) and sweet (4 mM AceK) stimuli.  $n = 18$  mice; data are mean  $\pm$  s.e.m. **b**, Schematic of optogenetic stimulation strategy; AAV-ChR2 was injected into the bitter cortex and the optical fibre was placed above the CEA. **c**, Quantification of licking response to water, bitter and sweet stimuli and sweet stimuli plus light.  $n = 6$  mice. Stimulation overrides the attractive responses to the sweet stimulus. **d**, Schematic of photostimulation of sweet terminals. **e**, Quantification of licking response to water, sweet and bitter stimuli and bitter stimuli plus light.  $n = 12$  mice. Stimulation overrides the aversive responses to the bitter stimulus. Repeated-measures one-way ANOVA followed by Bonferroni post hoc test; see Supplementary Table 1.

aversive response to an orally applied sweet tastant, whereas strong stimulation of sweet terminals in BLA might suppress aversion to an orally applied bitter tastant. We used a behavioural test in which thirsty mice expressing ChR2 in the bitter or sweet cortex were exposed to random presentations of water, a bitter chemical or a sweet solution (Fig. 4a). Next, we examined the effect of photoactivating bitter cortico-amygdalar projections by placing the stimulating optical fibre over the amygdala of mice that expressed ChR2 in the bitter cortex (Fig. 4b). Stimulation of bitter targets in the amygdala is indeed sufficient to transform the appetitive nature of a sweet tastant into an aversive one (Fig. 4c). Conversely, by photoactivating the amygdala targets of the sweet cortex it was possible to change the perceived valence of a bitter tastant (Fig. 4d, e). These results highlight the key role of the amygdala in imposing valence on a taste cue. To examine the effect of taste stimulation in the absence of amygdala function, we carried out a number of studies in which the neurons of the amygdala were reversibly silenced.

First, we used a behavioural assay that relies on direct stimulation of the taste cortex. In one group of mice, we introduced ChR2 into neurons in the sweet cortical field (Fig. 5a), bilaterally injected an AAV encoding inhibitory designer receptors exclusively activated by designer drugs (DREADD) into amygdala neurons for chemogenetic silencing<sup>22</sup> (see Methods for details), and tested the mice before and after clozapine *N*-oxide (CNO) injection. Importantly, ChR2 and the stimulating fibre are both in the sweet cortex (Fig. 5a), and because sweet neurons project to many targets (see Fig. 1a), the full repertoire is likely to be co-activated upon stimulation of the sweet cortical field. Notably, silencing of the amygdala was sufficient to abolish all attractive responses associated with activation of the sweet cortex (Fig. 5b); equivalent results were obtained using pharmacological inhibition of the amygdala with NBQX rather than inhibitory DREADD (Fig. 5c). We repeated similar studies but this time examined the activation of the bitter cortex (Fig. 5d). Our results showed that silencing of the amygdala is also sufficient to abolish aversive responses associated with the activation of the bitter cortex (Fig. 5e, f). Finally, we reasoned that the valence associated with sweet and bitter tastants delivered to the tongue, rather than direct stimulation of the taste cortex, could also be compromised. As predicted, the results shown in Fig. 5g–i demonstrate that silencing the amygdala impairs the behavioural preference for sweet chemicals and the aversion to bitter chemicals.

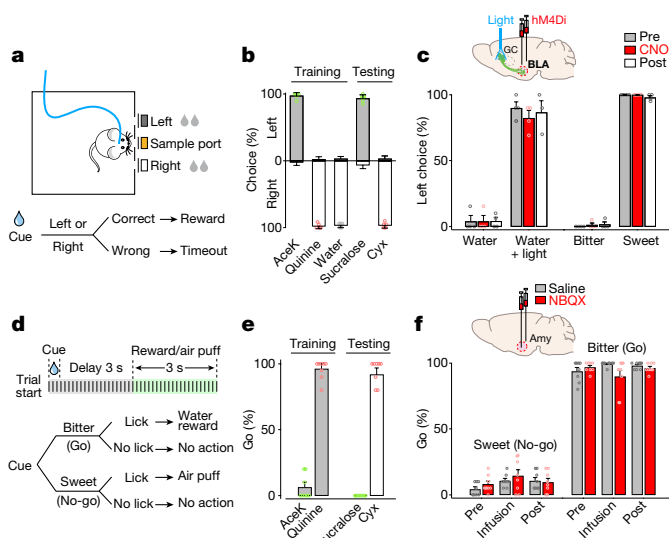


**Fig. 5 | Silencing neurons in the amygdala impairs taste valence.**

**a**, Schematic of optogenetic stimulation and the chemogenetic/pharmacological silencing strategy. AAV-ChR2 was injected unilaterally into the sweet cortex, and an optical fibre was implanted for photostimulation. AAV-hM4Di<sup>28</sup> was targeted bilaterally to the BLA for chemogenetic silencing (alternatively, cannulae were implanted bilaterally over the amygdala for pharmacological silencing). **b**, Chemogenetic silencing of the amygdala with inhibitory DREADD (hM4Di) and CNO abolished the strong appetitive behaviour observed following photostimulation of sweet cortex (compare pre and post with CNO).  $n = 4$  mice, water (W) versus water plus light (L), two-tailed paired *t*-test; pre,  $P = 0.0054$ ; CNO,  $P = 0.8900$ ; post,  $P = 0.0265$ . See Extended Data Fig. 5 for controls. **c**, Pharmacological silencing of the amygdala with NBQX similarly abolished the appetitive behaviour associated with photostimulation of the sweet cortex.  $n = 6$  mice, two-tailed paired *t*-test; pre,  $P = 0.0049$ ; NBQX,  $P = 0.9458$ ; post,  $P = 0.0042$ . See Extended Data Fig. 5 for controls. **d**, e, NBQX silencing of the amygdala abolished aversive responses to photostimulation of the bitter cortex.  $n = 3$  mice, two-tailed paired *t*-test; pre,  $P = 0.0047$ ; NBQX,  $P = 0.9125$ ; post,  $P = 0.0261$ . **f**, Saline controls for NBQX silencing following photostimulation of the bitter cortex.  $n = 3$  mice, two-tailed paired *t*-test; pre,  $P = 0.0261$ ; saline,  $P = 0.0230$ ; post,  $P = 0.0005$ . **g–i**, NBQX silencing of the amygdala diminished preference for sweet chemicals ( $n = 5$  mice) and aversion to bitter chemicals ( $n = 8$  mice); the small remaining responses to the orally applied sweet and bitter tastants probably reflect brain-stem-dependent immediate reactions to taste observed in decerebrated animals<sup>29</sup>. Repeated-measures one-way ANOVA followed by Bonferroni post hoc test (Supplementary Table 1). Values are mean  $\pm$  s.e.m.

Previously, we showed that silencing the sweet or bitter cortex prevented the recognition of sweet or bitter tastants, whereas optogenetic activation of those same cortical fields triggered prototypical sweet- and bitter-associated behaviours<sup>7</sup>. We reasoned that if tastant identity and valence are encoded in separate neural substrates, with the taste cortex responsible for imposing identity to a tastant, and the amygdala for affording its valence, then mice with silenced amygdala should still recognize the identity of a sweet or bitter taste stimulus, even if blind to its hedonic value.

We trained mice to report the identity of a tastant by using two different behavioural assays: a three-port test and a go/no-go assay. In the three-port test, mice learned to sample a taste cue from a centre spout (random presentations of water, a sweet or a bitter chemical), and then report its identity either by going to the right or left port;



**Fig. 6 | Silencing the amygdala does not prevent tastant recognition.** **a**, Schematic and flow chart for the three-port taste-recognition task. In each trial, a mouse had 0.5 s to lick a randomly presented taste cue from the middle port, and then go to either the left or right to report the identity of the tastant (in this example mice were trained to go left for sweet and right for bitter or water); correct responses were rewarded with 4 s of water, whereas incorrect ones led to a 5-s timeout penalty. **b**, Quantification of results from three-port recognition sessions, demonstrating highly reliable recognition of the stimulus identity (>90% accuracy).  $n = 6$  mice. Tastant concentrations: 10 mM AceK, 1 mM quinine, 3 mM sucralose, 10  $\mu$ M cycloheximide (Cyx). **c**, The mice used in Fig. 5b were assayed for the effect of silencing the amygdala.  $n = 4$  mice for pre, CNO;  $n = 3$  mice for post. Mice with a silenced amygdala can still identify the different tastes with normal accuracy. Importantly, photostimulation of the sweet cortex is recognized as a sweet-tasting stimulus<sup>7</sup>, and remains so after CNO silencing of the amygdala (compare water with water plus light). The graph only presents the responses to the left port (sweet identity). **d**, The animals used in Fig. 5e, h, i were assayed using go/no-go tastant recognition tests<sup>7</sup>. **e**, Mice show highly reliable recognition of the stimulus identity after training (>90% accuracy).  $n = 8$  mice. Tastant concentrations: 4 mM AceK, 1 mM quinine, 3 mM sucralose, 10  $\mu$ M Cyx. **f**, Pharmacological silencing of amygdala has no significant effect on either sweet or bitter recognition.  $n = 8$  mice. Two-way or repeated-measures one-way ANOVA followed by Bonferroni post hoc test (Supplementary Table 1). Values are mean  $\pm$  s.e.m.

a correct response was rewarded with 4 s of water (Fig. 6a). We initially focused on attractive responses as they represent the expression of a selective, positive behavioural response. After 20–30 sessions of training over 10–14 days (see Methods for details), trained mice were able to report the identity of each tastant in hundreds of randomized trials with over 90% accuracy (see Fig. 6b), showed correct behavioural responses to other sweet and bitter tastants that were not used in the training set (Fig. 6b, compare training to testing), and appropriately reported direct optogenetic activation of the sweet taste cortex as sweet, with approximately 90% of the water plus light trials producing correct responses (Fig. 6c, pre-silencing). Next, we assayed whether silencing of the amygdala using inhibitory DREADD in the BLA affected the ability of these mice to correctly identify a sweet stimulus. Our results (Fig. 6c) demonstrated that loss of amygdala function, while abolishing the ability of the sweet cortex to evoke appetitive responses (Fig. 5a–c), has no impact on the ability of the mice to properly identify sweet tastants (or to recognize light activation of the sweet cortex as sweet). As anticipated, inhibitory DREADD expression in the sweet cortex (just as has previously been shown using NBQX<sup>7</sup>), severely impairs sweet taste recognition (Extended Data Fig. 6).

The second behavioural platform relied on go/no-go behavioural assays, and examined both sweet and bitter recognition. Thirsty mice were trained to sample a test tastant from a spout, and then to report its

identity either by licking (go) or withholding licking (no-go; Fig. 6d). We trained animals to report 'go' when tasting bitter tastants and 'no-go' in response to sweet tastants, exactly the opposite of the innate drive. After 15–20 sessions of training, mice reported tastant identity with over 90% accuracy (Fig. 6e). Indeed, silencing of the amygdala, just as observed in the three-port tests, has no effect on recognition of sweet or bitter tastants (Fig. 6f). Notably, these experiments used the same mice that exhibited loss of sweet and bitter valence after NBQX infusion into the amygdala (Fig. 5). Taken together, these studies show that the amygdala is necessary and sufficient to drive valence-specific behaviours to taste stimuli and that the cortex can independently represent taste identity.

The senses of taste and smell function as the principal gateways for assessing the attraction to, and palatability of, food cues. In its most fundamental state, taste mediates innate consummatory and rejection behaviours, while also allowing an animal to learn the association of food sources with hardwired tastant-dependent actions. Here we studied the neural basis for innate responses to sweet and bitter, and showed that the taste cortex and the amygdala function as two essential, but distinct, neural stations for identifying tastants and for imposing valence on sweet and bitter.

Recent molecular studies have identified distinct populations of neurons in the amygdala that may serve as neural substrates for a wide range of positive and negative hedonic responses<sup>10,13–19</sup>. In this study, we show that sweet and bitter cortical fields exhibit separate projection targets in the amygdala, and that photoactivation of these cortico-amygdalar projections evokes opposing responses. However, these can be experimentally dissociated from the cortex, such that animals may recognize a 'taste stimulus' but remain oblivious to its valence. Together, these results provide an anatomical substrate for imposing hedonic value to sweet and bitter, and the basic logic for the generation of hardwired, stereotypic attractive and aversive taste responses.

The amygdala is known to provide representations of Pavlovian associations<sup>11,23,24</sup>, such that innately rewarding and aversive tastants may also function as unconditioned stimuli in conditioning protocols<sup>25</sup>. Therefore, in addition to imposing valence on tastants, the amygdala probably links taste valence to other stimuli so that associative memories can be formed, and thereby appropriate valence-specific behaviour may be elicited by previously neutral cues from other modalities that would now predict a bitter or sweet tastant. Notably, the sweet and bitter cortex project to several additional brain areas, including those involved in feeding, motor systems, multisensory integration, learning and memory (Fig. 1). In the future, it will be exciting to unravel how these circuits come together to drive innate and learned responses.

## Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0165-4>.

Received: 3 October 2017; Accepted: 26 March 2018;

Published online: 30 May 2018

- Yarmolinsky, D. A., Zuker, C. S. & Ryba, N. J. Common sense about taste: from mammals to insects. *Cell* **139**, 234–244 (2009).
- Scott, K. Taste recognition: food for thought. *Neuron* **48**, 455–464 (2005).
- Spector, A. C. & Travers, S. P. The representation of taste quality in the mammalian nervous system. *Behav. Cogn. Neurosci. Rev.* **4**, 143–191 (2005).
- Accolla, R., Bathellier, B., Petersen, C. C. & Carleton, A. Differential spatial representation of taste modalities in the rat gustatory cortex. *J. Neurosci.* **27**, 1396–1404 (2007).
- Yoshimura, H., Sugai, T., Fukuda, M., Segami, N. & Onoda, N. Cortical spatial aspects of optical intrinsic signals in response to sucrose and NaCl stimuli. *Neuroreport* **15**, 17–20 (2004).
- Chen, X., Gabitto, M., Peng, Y., Ryba, N. J. & Zuker, C. S. A gustotopic map of taste qualities in the mammalian brain. *Science* **333**, 1262–1266 (2011).
- Peng, Y. et al. Sweet and bitter taste in the brain of awake behaving animals. *Nature* **527**, 512–515 (2015).
- Susaki, E. A. et al. Whole-brain imaging with single-cell resolution using chemical cocktails and computational analysis. *Cell* **157**, 726–739 (2014).

9. Zingg, B. et al. AAV-mediated anterograde transsynaptic tagging: mapping corticocollicular input-defined neural pathways for defense behaviors. *Neuron* **93**, 33–47 (2017).
10. Cai, H., Haubensak, W., Anthony, T. E. & Anderson, D. J. Central amygdala PKC- $\delta^{+}$  neurons mediate the influence of multiple anorexigenic signals. *Nat. Neurosci.* **17**, 1240–1248 (2014).
11. Janak, P. H. & Tye, K. M. From circuits to behaviour in the amygdala. *Nature* **517**, 284–292 (2015).
12. Paton, J. J., Belova, M. A., Morrison, S. E. & Salzman, C. D. The primate amygdala represents the positive and negative value of visual stimuli during learning. *Nature* **439**, 865–870 (2006).
13. Douglass, A. M. et al. Central amygdala circuits modulate food consumption through a positive-valence mechanism. *Nat. Neurosci.* **20**, 1384–1394 (2017).
14. Kim, J., Pignatelli, M., Xu, S., Itohara, S. & Tonegawa, S. Antagonistic negative and positive neurons of the basolateral amygdala. *Nat. Neurosci.* **19**, 1636–1646 (2016).
15. Namburi, P. et al. A circuit mechanism for differentiating positive and negative associations. *Nature* **520**, 675–678 (2015).
16. Gore, F. et al. Neural representations of unconditioned stimuli in basolateral amygdala mediate innate and learned responses. *Cell* **162**, 134–145 (2015).
17. Han, W. et al. Integrated control of predatory hunting by the central nucleus of the amygdala. *Cell* **168**, 311–324 (2017).
18. Kim, J., Zhang, X., Muralidhar, S., LeBlanc, S. A. & Tonegawa, S. Basolateral to central amygdala neural circuits for appetitive behaviors. *Neuron* **93**, 1464–1479 (2017).
19. Beyeler, A. C. et al. Organization of valence-encoding and projection-defined neurons in the basolateral amygdala. *Cell Rep.* **22**, 905–918 (2018).
20. Boyden, E. S., Zhang, F., Bamberg, E., Nagel, G. & Deisseroth, K. Millisecond-timescale, genetically targeted optical control of neural activity. *Nat. Neurosci.* **8**, 1263–1268 (2005).
21. Tye, K. M. et al. Amygdala circuitry mediating reversible and bidirectional control of anxiety. *Nature* **471**, 358–362 (2011).
22. Stachniak, T. J., Ghosh, A. & Sternson, S. M. Chemogenetic synaptic silencing of neural circuits localizes a hypothalamus→midbrain pathway for feeding behavior. *Neuron* **82**, 797–808 (2014).
23. Maren, S. Neurobiology of Pavlovian fear conditioning. *Annu. Rev. Neurosci.* **24**, 897–931 (2001).
24. Cardinal, R. N., Parkinson, J. A., Hall, J. & Everitt, B. J. Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. *Neurosci. Biobehav. Rev.* **26**, 321–352 (2002).
25. Schoenbaum, G., Chiba, A. A. & Gallagher, M. Neural encoding in orbitofrontal cortex and basolateral amygdala during olfactory discrimination learning. *J. Neurosci.* **19**, 1876–1884 (1999).
26. Lein, E. S. et al. Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445**, 168–176 (2007).
27. Bakker, R., Tiesinga, P. & Kötter, R. The scalable brain atlas: instant web-based access to public brain atlases and related content. *Neuroinformatics* **13**, 353–366 (2015).
28. Armbruster, B. N., Li, X., Pausch, M. H., Herlitze, S. & Roth, B. L. Evolving the lock to fit the key to create a family of G protein-coupled receptors potentially activated by an inert ligand. *Proc. Natl Acad. Sci. USA* **104**, 5163–5168 (2007).
29. Grill, H. J. & Norgren, R. The taste reactivity test. II. Mimetic responses to gustatory stimuli in chronic thalamic and chronic decerebrate rats. *Brain Res.* **143**, 281–297 (1978).

**Acknowledgements** We thank M. Tessier-Lavigne, N. Renier and P. Ariel for help with CUBIC; members of the Zuker laboratory and R. Axel for helpful discussions. We also acknowledge the Bio-Imaging Resource Center at Rockefeller University. Research reported in this publication was supported by the National Institute on Drug Abuse of the National Institutes of Health under award number R01DA035025 (C.S.Z.). N.J.P.R. is supported by the Intramural Research Program of the NIH, NIDCR; and C.D.S. was supported by R01 MH082017 from NIMH. C.S.Z. is an investigator of the Howard Hughes Medical Institute and a Senior Fellow at Janelia Farm Research Campus.

**Reviewer information** *Nature* thanks I. de Araujo and P. Kenny for their contribution to the peer review of this work.

**Author contributions** L.W. and S.G.-S. designed the study, carried out the experiments and analysed data, Y.P. and J.Z. performed behavioural experiments. C.S.Z., X.C., N.J.P.R. and C.D.S. designed the study and analysed data. C.S.Z., L.W. and N.J.P.R. wrote the paper.

**Competing interests** The authors declare no competing interests.

#### Additional information

**Extended data** is available for this paper at <https://doi.org/10.1038/s41586-018-0165-4>.

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41586-018-0165-4>.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>.

**Correspondence and requests for materials** should be addressed to C.S.Z.  
**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## METHODS

**Animals and surgery procedures.** All procedures were carried out in accordance with the US National Institutes of Health (NIH) guidelines for the care and use of laboratory animals, and were approved by the Columbia University Institutional Animal Care and Use Committee. Seven- to nine-week-old male C57BL/6J mice and B6.Cg-Gt(ROSA)<sup>tm9(CAG-tdTomato)Hze/J</sup> mice (Ai9)<sup>30</sup> were used for viral injections.

Animals were anaesthetized with ketamine and xylazine (100 mg kg<sup>-1</sup> and 10 mg kg<sup>-1</sup>, intraperitoneal), placed into a stereotaxic frame with a close-loop heating system to maintain body temperature, and unilaterally injected with 20–50 nl of AAV carrying ChR2 (AAV9-CamKIIα-hChR2(H134R)-EYFP-WPRE-SV40, Penn Vector Core) either in the sweet cortical field (bregma 1.7 mm; lateral 3.1 mm; ventral 1.8 mm) or the bitter cortical field (bregma -0.35 mm; lateral 4.2 mm; ventral 2.7 mm). The location of the taste cortex was verified by anatomical and optogenetic assays. Anterograde tracing<sup>6</sup> and retrograde tracing (Extended Data Fig. 7) showed that these cortical areas receive input from the taste thalamus (VPMpc). Photostimulation of these sweet and bitter cortical fields evokes prototypical attractive and aversive taste behaviours, respectively<sup>7</sup>. We also examined behavioural data from 14 ChR2 injections in the middle (Extended Data Fig. 8); six of the animals showed a modest increase in lick responses, 3 exhibited no change and 5 showed a small range of aversion. We believe this variability probably reflects the spread of the injection site.

Following viral injections, a customized implantable fibre (core diameter 200 μm, NA 0.39) was implanted 300–500 μm above the injection site, and guide cannulas (26 gauge, PlasticsOne) were unilaterally or bilaterally implanted above the anterior BLA (bregma -1.0 mm; lateral 3.2 mm; ventral 3.7 mm) or CEA (bregma -1.2 mm; lateral 3.0 mm; ventral 3.7 mm). These guide cannulas were used both for photostimulation of cortical projections and intracranial infusion in pharmacological silencing experiments. A metal head post was attached for head fixation during behavioural tests. All implants were secured onto the skull with dental cement (Lang Dental Manufacturing). For chemogenetic silencing experiments, 150–250 nl of AAV carrying hM4Di (AAV8-hSyn-hM4Di-mCherry, UNC Vector Core) was injected bilaterally into the BLA (bregma -1.0 mm; lateral 3.2 mm; ventral 4.2 mm) or sweet cortical field (bregma 1.7 mm; lateral 3.1 mm; ventral 1.8 mm) at a slow rate (15 nl min<sup>-1</sup>). All ventral coordinates listed above are relative to the pial surface. Mice were allowed to recover for at least 2–3 weeks before the start of behavioural experiments. For anterograde transsynaptic tracing, AAV1-hSyn-Cre<sup>9</sup> (20–50 nl) was injected into the sweet or bitter cortical field of mice carrying a Cre-dependent tdTomato reporter (Ai9<sup>30</sup>). Mice were examined four weeks after the injection. Placements of viral injections, guide cannulas and implanted fibres were histologically verified at the termination of the experiments using DAPI (1:5,000, Thermo Fisher Scientific) or TO-PRO-3 (1:1,000, Thermo Fisher Scientific) staining of 100-μm coronal sections. A confocal microscope (FV1000, Olympus) was used for fluorescence imaging.

**Whole-brain clearing and imaging.** For whole-brain tracing of the projections of cortical neurons, we unilaterally injected a small volume (10–20 nl) of mixed AAVs carrying Cre-recombinase and Cre-dependent eGFP (AAV1-CamKII0.4-Cre-SV40 and AAV1-CAG-Flex-eGFP-WPRE-bGH, 1:100, Penn Vector Core) in the sweet cortical field, and the same volume of mixed AAVs carrying Cre-recombinase and Cre-dependent tdTomato (AAV1-CamKII0.4-Cre-SV40 and AAV9-CAG-Flex-tdTomato-WPRE-bGH, 1:100, Penn Vector Core) in the bitter cortical field. Four weeks after AAV injection, mice were transcardially perfused with 5–10 ml of phosphate-buffered saline (PBS) containing 10 U ml<sup>-1</sup> heparin, followed by 20 ml 4% paraformaldehyde; brains were post-fixed in 4% paraformaldehyde for an additional 3 h at room temperature. Whole brains were then treated following the CUBIC clearing protocol<sup>8,31</sup>. To prevent sample deformation caused by temperature fluctuation and to minimize fluorescence loss during clearing, all clearing procedures were performed at room temperature. CUBIC clearing reagents were prepared as previously described<sup>8,31</sup>. Reagent 1 contained 25 wt% urea (Sigma-Aldrich), 25 wt% N,N,N',N'-tetrakis(2-hydroxypropyl)ethylenediamine (Sigma-Aldrich) and 15 wt% Triton X-100 (Nacalai Tesque). Reagent 2 contained 50 wt% sucrose (Sigma-Aldrich), 25 wt% urea, 10 wt% triethanolamine (Sigma-Aldrich) and 0.1% (v/v) Triton X-100. The fixed brains were washed three times with PBS, immersed in reagent 1 (diluted 1:2 in water) overnight with gentle shaking and then incubated in reagent 1 for 7–10 days with gentle shaking. Brains were washed with PBS, degassed in PBS overnight and were then transferred into 5 ml of reagent 2 diluted 1:2 in PBS for 6–24 h before immersion in reagent 2 for 3–7 days for further clearing and reflection index matching. TO-PRO-3 (1:5,000, Thermo Fisher Scientific) was added to reagent 2 for counterstaining. During the immersion in reagent 2, tubes were not shaken to avoid bubbles. Samples were kept in reagent 2 for up to one week at room temperature before imaging.

On the day of imaging, samples were gently wiped to remove reagent 2 residue and transferred into an oil mix (mineral oil and silicone oil 1:1, final refraction index 1.48–1.49) at least 1 h before imaging. Light-sheet fluorescence microscopy

(UltraMicroscope, LaVision BioTec) with a 2× objective lens (0.5 NA, working distance 10 mm) or 4× objective lens (0.3 NA, working distance 6 mm) was used for rapid image acquisition of the whole brain. The samples were sequentially illuminated with a unidirectional light sheet produced by 488-nm, 561-nm and 640-nm lasers and scanned with a z-step size of 8.13–13 μm from ventral to dorsal. Exposure time was 50–200 ms per channel per z step. To cover the whole brain, each sample was imaged either via 4 × 4 tile scans with the 4× lens or using multi-position scans with the 2× lens (three manually assigned positions to cover two hemispheres and brainstem).

Whole-brain image tiles were scaled to 1/8 of the original size and stitched in three dimensions using ImageJ 1.51n (Fiji distribution). The 640-nm channel of the whole-brain data was registered to a reference atlas (Allen Brain Institute, 25-μm resolution volumetric data with annotation map, <http://www.brain-map.org>) using ANTs (advanced normalization tools 1.9.x) with affine transformation<sup>26,32</sup>. The same transformation was applied to the other two channels using the WarpImageMultiTransform function of ANTs. z projections of maximum intensity and virtual sections were processed in ImageJ (noise was filtered with the remove outlier function). Because of the high dynamic range of the fluorescent intensity between the soma and the fine processes, the gamma value of the images shown in Fig. 1 was set to 0.5 for display purposes.

**Fos induction and immunohistochemistry.** Mice expressing ChR2 in the sweet or bitter cortex were habituated by performing mock stimulations (see below) once a day for three days before Fos induction. On the day of the experiment, mice were photostimulated for 30 min (473 nm, 20 Hz, 20-ms pulses, 5 s on and 5 s off, 5–10 mW per mm<sup>2</sup>). Mice were then allowed to rest for 1 h and were processed for immunostaining as previously described<sup>7</sup>. Tissue sections were incubated with goat anti-Fos antibody (1:500, Santa Cruz, sc-52-G) for 24 h at 4 °C. Fluorescently tagged secondary antibodies (Alexa-594 donkey anti-goat or Alexa-647 donkey anti-goat, 1:1,000, Thermo Fisher Scientific) were used to visualize Fos expression. All sections were imaged using an Olympus FV-1000 confocal microscope.

**Head-restrained lick preference assays.** Head-restrained lick preference assays were performed as previously described<sup>7</sup>. Mice expressing ChR2 in the sweet or bitter cortex were initially water-deprived for 24 h to motivate drinking in head-restrained assays and then acclimated to drinking from a motor-positioned spout (two sessions per day for at least two days) before testing. Mice were weighed daily during the behavioural assays and supplied with necessary water to maintain at least 85% of their initial body weight. Each trial began with a light cue, followed 1 s later by the spout swinging into position and a tone cue to indicate the onset of tastant delivery; after 5 s (during which the mouse could lick) the spout rotated out of position. To measure attractive responses, mice were mildly water restricted (water-deprived for 24 h, and then provided with water until they exhibited an average of 5–15 licks per 5-s window); mice were supplied with 2–5 μl water at the beginning of each trial. To measure aversion, mice were water-deprived for 24 h, and supplied with 5–10 μl water per trial; mice normally exhibited active licking over the full five seconds (average 30–40 licks per 5 s as a sign of thirst). Training sessions consisted of 60 trials with water; testing sessions shown in Fig. 3 consisted of 15 trials with water, 4 of which were coupled to photostimulation of cortical terminals in the amygdala; testing sessions in Fig. 5b–f consisted of 20 trials with water, 10 of which were pseudo-randomly coupled to photostimulation of the sweet cortex; testing sessions in Fig. 4 consisted of 60 trials, 20 trials with water, 20 trials with bitter taste (0.5 mM quinine), 20 trials with sweet taste (4 mM AceK). To examine the effect of amygdalar nuclei on taste preference, 50% of sweet trials in Fig. 4c or 50% of bitter trials in Fig. 4e were pseudo-randomly coupled to photostimulation of the CEA (Fig. 4c) or BLA (Fig. 4e). The delivery of tastants was triggered by licking actions such that mice could consume as much or little as they chose during the 5 s. To minimize the influence of thirst and satiety on the assessment of taste palatability, for each test session in Fig. 4 we included consecutive trials satisfying two criteria: (1) licks to bitter less than 20 (otherwise mice were too thirsty); (2) more than 5 licks to water (otherwise mice were already satiated). The licking behaviour was videotaped during the entire session and licking events were identified by a custom-written code in MATLAB. For photostimulation, 473-nm light stimuli (diode-pumped solid-state laser, Shanghai Laser & Optics Century Co. or fibre-coupled LED, Thorlabs) were delivered via an optical fibre inserted into a guide cannula over the amygdala or via an implantable fibre over the taste cortex. Light stimulation was controlled by contact of the tongue with the metal spout; one lick triggered a train of light pulses (10–20 Hz, 20 ms per pulse, 20 pulses, 5–15 mW per mm<sup>2</sup>). Licks during the light stimulation extended the stimulus until 1 s after the last lick. Light/tone cues, the delivery of tastants and light stimuli were controlled using a MATLAB program via a microcontroller board (Arduino Mega 2560, Arduino)<sup>7</sup>. Each point in Fig. 3c–f and Fig. 4 indicates data averaged from multiple test sessions for an individual mouse. In Fig. 3e, f, the lick ratio refers to the number of licks in the presence of light stimulation over the number of licks in water-only trials.

**Free-moving lick preference assays.** Taste preference (Fig. 5g–i) was also measured in free-moving animals by using a custom-built gustometer<sup>33</sup>. Prior to testing, mice were water-restricted for 24 h, and then provided with water until they exhibited an average of <20 licks per 5-s window (to test attractive responses). Alternatively, after 24 h of water-restriction mice were provided with unrestricted water access for 5–10 min, and then assayed 18 h later (that is, to test aversive responses animals need to be sufficiently thirsty to be motivated to sample an unattractive cue). For testing, mice were presented with water versus 4 mM AceK or water versus 1 mM quinine as previously described<sup>33</sup>.

**Place-preference assays.** Mice expressing Chr2 in sweet or bitter cortex were tested in a custom-built two-chamber arena placed inside a sound-attenuating cubicle; the arena (30 cm × 15 cm), was designed with two chambers, one with alternating black and white vertical stripes, and the other with alternating black and white squares. Animal locations were tracked in real-time by videotaping<sup>7</sup>. Mice were tested in the arena for 30 min with photostimulation of the sweet or bitter cortico-amygdalar projections via an optical fibre above the BLA or CEA, respectively. The last 15 min of each testing session was used to calculate the preference index (PI);  $PI = (t_1 - t_2) / (t_1 + t_2)$ , where  $t_1$  is the fraction of time a mouse spent in the chamber 1 (stimulating chamber), and  $t_2$  is the time spent in chamber 2 (non-stimulating chamber). For photostimulation of the sweet cortico-amygdalar projections, light was delivered for 5 s, with a 3-s interval (20 Hz, 20-ms pulses, 5–10 mW per mm<sup>2</sup>) to avoid over-stimulation or phototoxicity; for photostimulation of the bitter cortico-amygdalar projection, light (20 Hz, 20-ms pulses, 2–5 mW per mm<sup>2</sup>) was delivered for 1 s with a 3-s interval; a sound cue was used to mark the onset of each stimulation<sup>15,34,35</sup>. For each cohort (8 mice for sweet and 5 animals for bitter), half the animals were tested with light on in the baseline-preferred chamber, and half with light on in the baseline unpreferred chamber<sup>15,34</sup>. When the mouse crossed to the non-stimulating chamber, the light was automatically turned off immediately.

**Three-port taste-recognition assays.** Mice deprived of water for 24 h were trained to perform a taste-recognition task in a customized three-port behaviour chamber in which they sampled taste cues from the middle port and then reported the taste identity of the cue by choosing to lick from either the left or right port. Taste cues (AceK, quinine or water) were pseudo-randomly delivered through a metal spout in the middle port. Each trial began with the shutter opening in the middle port. Mice were given (up to) 15 s to initiate a trial by licking the middle spout (failure to initiate a trial resulted in the shutter closing and a new trial starting). The shutter in the middle port closed 0.5 s after the first lick allowing animals 0.5 s to sample tastants cues (2–3  $\mu$ l); 0.5 s after the middle port closed, the shutters of the left and right ports opened simultaneously. Mice were given 4 s to make a left or right choice and obtain the water reward (total ~6  $\mu$ l). For a given mouse, reward from side ports was assigned to taste cues (for example, left for sweet, right for bitter and water). A wrong choice triggered a penalty of a 5-s timeout. The inter-trial interval was 1 s. Mice were trained for two sessions per day, with 80–100 pseudo-randomized trials per session until they could effectively discriminate the tastants with approximately 90% accuracy (2–3 weeks). To test the effect of photostimulation of sweet cortical neurons, mice expressing Chr2 in the sweet cortex were trained to discriminate sweet from bitter and water (for example, left for sweet, right for either bitter or water) and then tested with sweet, bitter, water and water with light (473 nm, 20 Hz, 20 ms per pulse, 20 pulses triggered by one lick of the middle spout). A testing session consisted of 20 sweet trials, 20 bitter trials, 10 water-only trials and 10 water with light trials. To avoid mice using photostimulation light as a visual cue, the connection between implantable fibre and patch cable was properly shielded. To prevent learning during the test, no time-out penalties were given and no reward was provided for water with light trials. Performances were calculated as the percentage of correct choices for a given taste cue. The lick behaviour was detected by a capacitive touch sensor (MPR121, SparkFun). The delivery of tastants, shutter position and light stimuli were controlled by a custom-written program in MATLAB via an Arduino board.

**Go/no-go taste-recognition assays.** Go/No-go taste-recognition assays were performed as previously described<sup>7</sup>. Mice were trained until they could effectively discriminate the tastants with approximately 90% accuracy (over 1–2 weeks).

On the ‘probe’ sessions, no punishment was applied for ‘no-go’ tastants to avoid re-learning; neither reward nor punishment were delivered to novel tastants.

**Pharmacological inhibition.** The selective AMPA receptor antagonist NBQX (2,3-dioxo-6-nitro-1,2,3,4-tetrahydrobenzo[f]quinoxaline-7-sulfonamide, 5 mg ml<sup>-1</sup> in 0.9% NaCl, 100–300 nl, Tocris Bioscience) was unilaterally (Fig. 3e, f) or bilaterally (Figs. 5, 6) infused into the amygdala using a 1- $\mu$ l microsyringe (Hamilton) and an internal cannula (PlasticsOne) inserted into the guide cannula above the amygdala. The infusion rate was approximately 100 nl min<sup>-1</sup>. After intracranial infusion, mice were allowed to rest in their home cage for 1–1.5 h before re-test. A post-test was performed after a recovery period of 24 h. As a control, the same experiment was conducted using isotonic saline (0.9% NaCl) in the same animals.

**Chemogenetic inhibition.** Mice injected with Chr2 in the sweet cortex and hM4Di in the BLA were first tested in the head-restrained lick preference assay as described above. To effectively examine inhibition using hM4Di, we determined (and used) the minimum light intensity for photostimulation that produced significant attractive responses (pre-test). On the following day, CNO was injected (10 mg kg<sup>-1</sup>, intraperitoneal) and the behavioural test with the same level of photostimulation was repeated between 1 and 2 h after CNO injection. Mice were allowed to rest and recover in their home cage and a post-test was performed at least 24 h later.

The same mice were then trained in the three-port assay for taste recognition. After achieving at least 90% accuracy in training sessions, mice were tested for the ability to recognize tastants and optogenetic stimulation of the sweet cortex in the three-port taste-recognition assay before and after chemogenetic silencing (pre: 24 h before CNO injection; CNO: between 1–2 h after injection; post: at least 24 h after injection). To confirm that amygdala was indeed efficiently silenced in the experiments presented in Fig. 6c, mice were tested for attractive responses to photostimulation of the sweet cortex in a lick preference assay; this was performed after each three-port session before and after chemogenetic silencing.

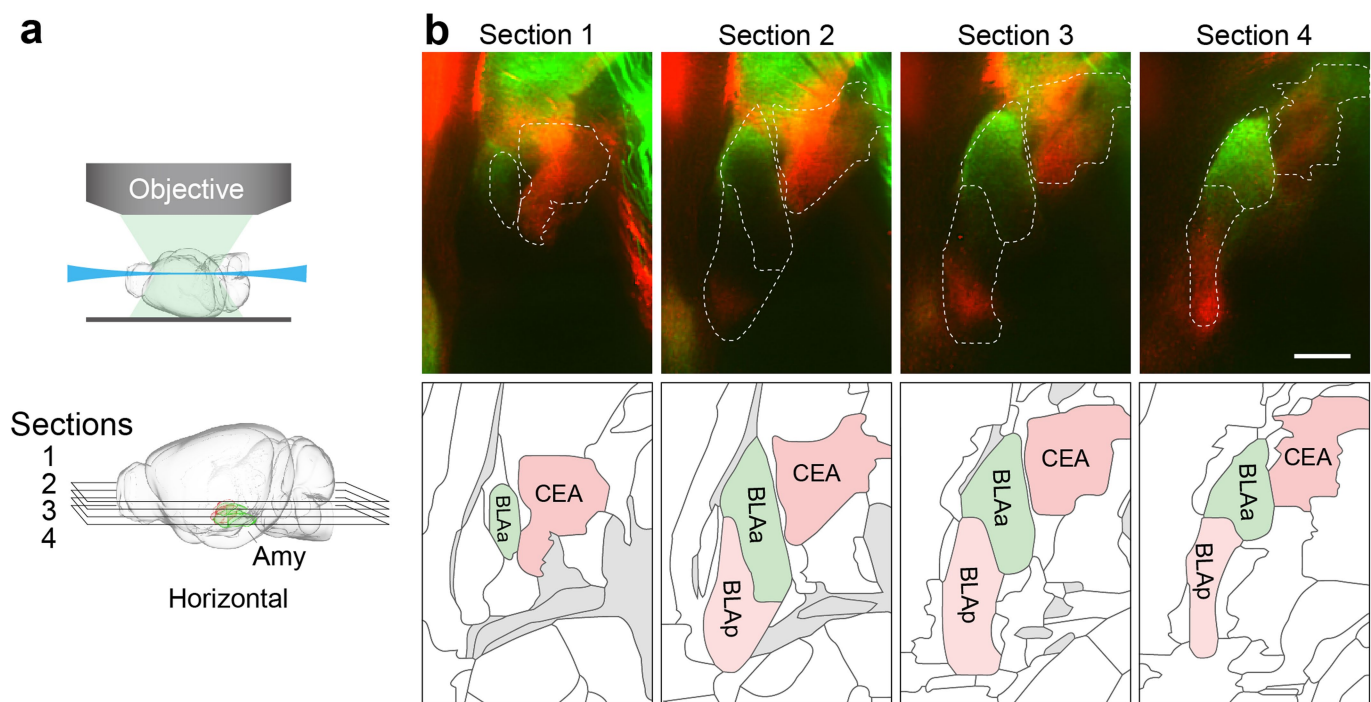
**Statistics.** No statistical methods were used to predetermine sample size, and investigators were not blinded to group allocation. No method of randomization was used to determine how animals were allocated to experimental groups. Animals in which post hoc histological examination showed that viral targeting or the position of implanted fibre/cannulas was in the wrong location were excluded from analysis. Statistical methods are indicated when used, and statistical analyses for all figures are provided in Supplementary Table 1. Multiple comparisons were analysed using repeated-measures one-way or two-way ANOVAs followed by the Bonferroni correction. All analyses were performed in MATLAB 2016a (MathWorks), Prism 7.0a (GraphPad) and Igor Pro 6.37 (WaveMetrics). Data are presented as mean  $\pm$  s.e.m.

**Reporting summary.** Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

**Code availability.** Custom code for behavioural assays is available from the corresponding author upon reasonable request.

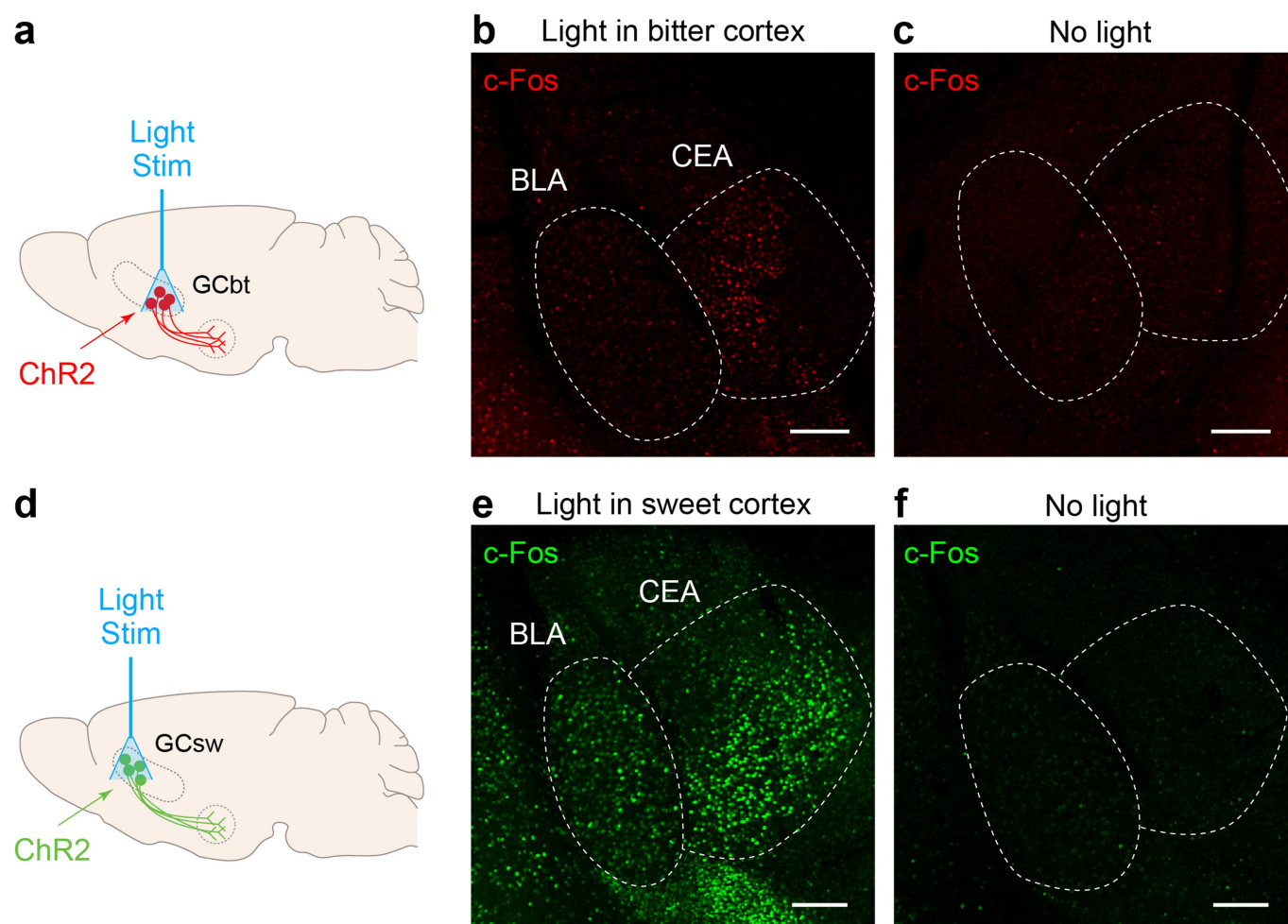
**Data availability.** All data supporting the findings of this study are available from the corresponding author upon reasonable request.

30. Madisen, L. et al. A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nat. Neurosci.* **13**, 133–140 (2010).
31. Susaki, E. A. et al. Advanced CUBIC protocols for whole-brain and whole-body clearing and imaging. *Nat. Protoc.* **10**, 1709–1727 (2015).
32. Avants, B. B. et al. A reproducible evaluation of ANTs similarity metric performance in brain image registration. *Neuroimage* **54**, 2033–2044 (2011).
33. Zhang, Y. et al. Coding of sweet, bitter, and umami tastes: different receptor cells sharing similar signaling pathways. *Cell* **112**, 293–301 (2003).
34. Kim, C. K. et al. Molecular and circuit-dynamical identification of top-down neural mechanisms for restraint of reward seeking. *Cell* **170**, 1013–1027 (2017).
35. Ilango, A. et al. Similar roles of substantia nigra and ventral tegmental dopamine neurons in reward and aversion. *J. Neurosci.* **34**, 817–822 (2014).
36. Krettek, J. E. & Price, J. L. A description of the amygdaloid complex in the rat and cat with observations on intra-amygdaloid axonal connections. *J. Comp. Neurol.* **178**, 255–279 (1978).



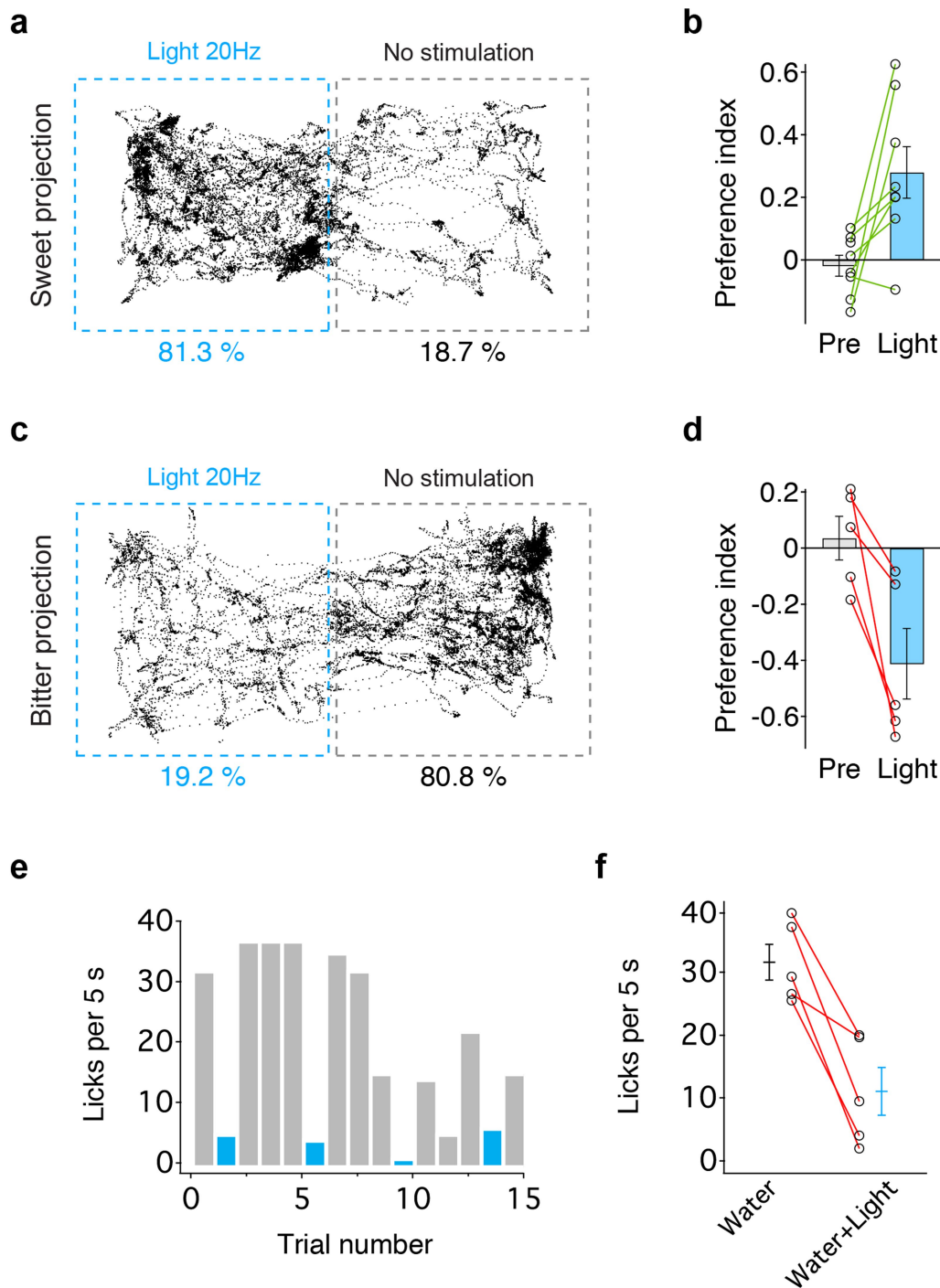
**Extended Data Fig. 1 | Projections from the sweet and the bitter cortex terminate in distinct targets in the amygdala.** **a**, The cartoon illustrates the imaging planes of the optical horizontal sections at different depths of the amygdala. The brain diagrams were rendered by the scalable brain composer ([https://scalablebrainatlas.incf.org/services/sba-composer.php?template=ABA\\_v3](https://scalablebrainatlas.incf.org/services/sba-composer.php?template=ABA_v3)) based on the Allen Mouse Brain Common Coordinate Framework version 3<sup>26,27</sup>. **b**, Segregation of sweet and bitter cortical projections in the amygdala. Sweet cortical neurons project to

anterior BLA (green), whereas bitter cortical neurons predominantly innervate the CEA (red) and a portion of the posterior BLA (red). Top, optical horizontal sections at different dorsal–ventral positions are shown (sections 1–4; see **a**). Bottom, the boundaries of amygdala nuclei were determined by aligning fluorescence images to the Allen Brain Institute atlas<sup>26</sup> (<http://brain-map.org/>). Scale bar, 500  $\mu\text{m}$ . Similar results were observed in six independently labelled and imaged animals.



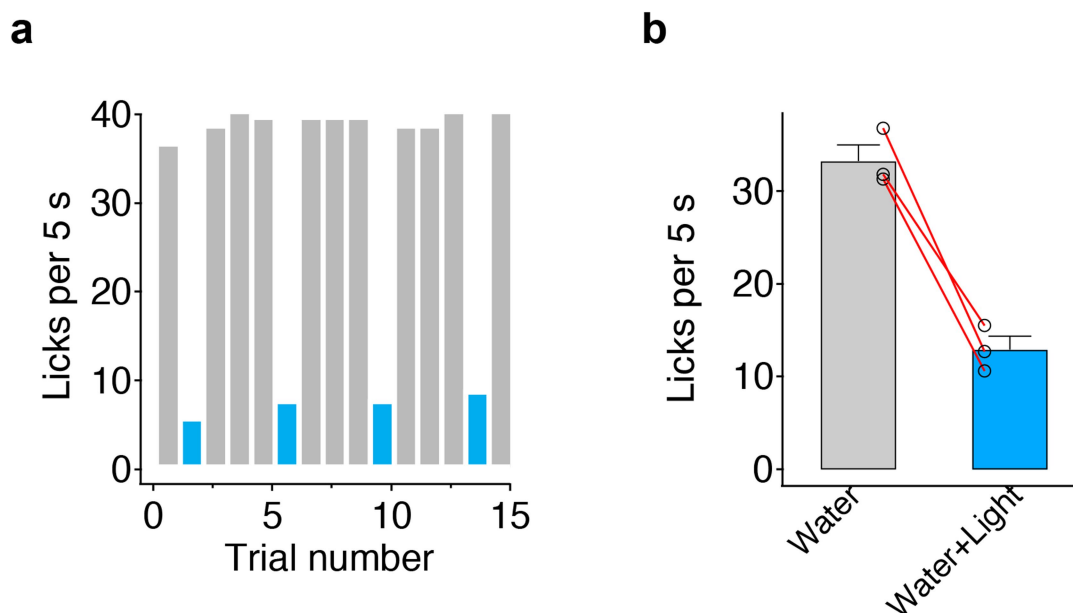
**Extended Data Fig. 2 | Activity-dependent labelling of sweet and bitter cortical targets in the amygdala.** Fos expression<sup>16</sup> in response to optogenetic activation of the sweet and bitter cortical fields was used to label activated neurons. **a**, Schematic of optogenetic stimulation strategy in the bitter cortex for Fos induction. **b**, Fos expression in the amygdala in response to photostimulation of the bitter cortex. The majority of Fos<sup>+</sup>

neurons are localized in the CEA. **c**, Fos expression in a control mouse without light stimulation. **d–f**, Photostimulation of the sweet cortex induces Fos expression in the amygdala. Note that the CEA, as a major local output of the BLA<sup>11,36</sup>, also shows strong Fos labelling in response to photostimulation. Scale bars, 200  $\mu$ m. Similar results were obtained in three animals for each experiment.



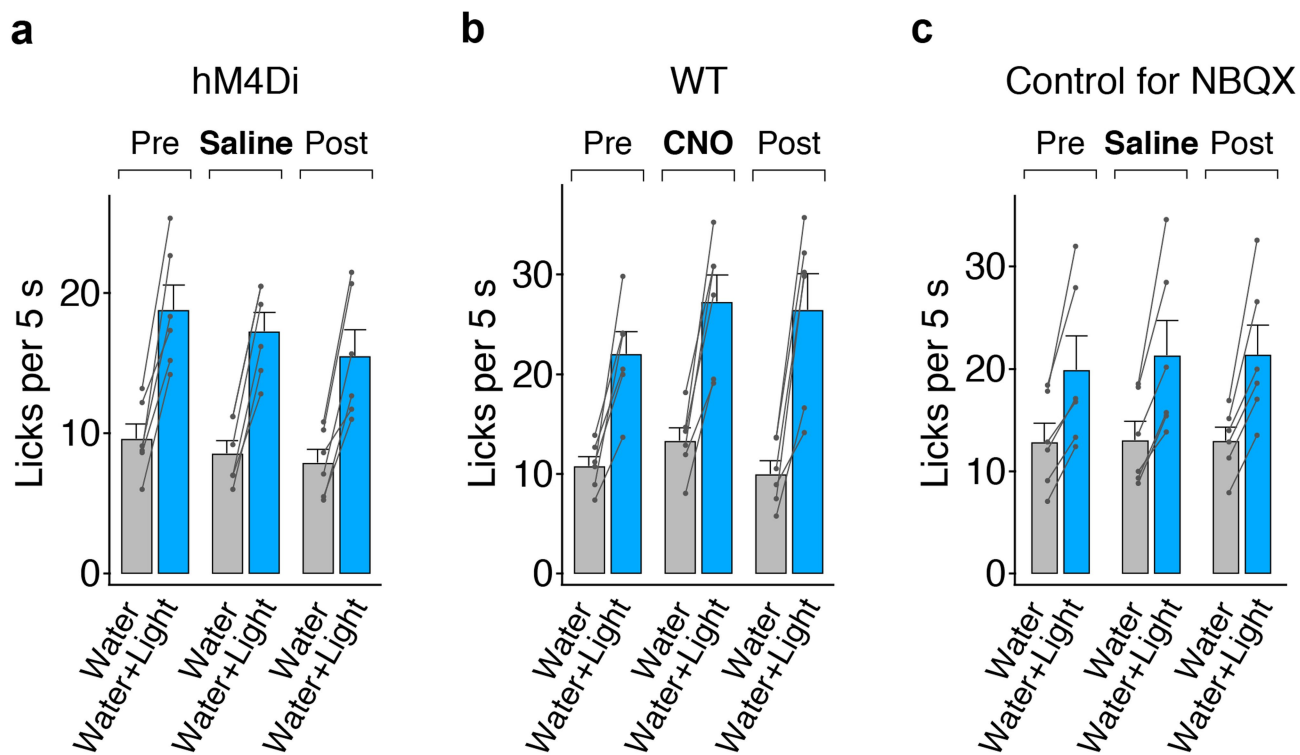
**Extended Data Fig. 3 | Place preference by photostimulation of cortico-amygdalar projections.** **a**, Representative tracking of a mouse during the 15-min place-preference test in a two-chamber arena; the left chamber in the diagram was coupled to stimulation of sweet cortico-amygdalar projections (see Methods for details); this animal spent over 80% of the test time in the chamber linked to stimulation of sweet projections to the amygdala. **b**, Quantification of preference index before (Pre) and during light stimulation.  $n = 8$  mice, two-tailed paired  $t$ -test,  $P = 0.0156$ . **c**, **d**, Place-preference test with stimulation of bitter cortical projections in the amygdala.  $n = 5$  mice, two-tailed paired  $t$ -test,  $P = 0.0207$ . **e**, **f**, Animals

used in **d** were also tested in a licking assay with similar light stimulation intensity, demonstrating strong suppression of licking responses.  $n = 5$  mice, two-tailed paired  $t$ -test,  $P = 0.0056$ . Values are mean  $\pm$  s.e.m. We note that we have examined multiple independent behavioural experiments activating sweet projections to the BLA and have never observed the induction of motor patterns or consummatory behaviour. Strong stimulation of bitter cortico-amygdalar projections (20 Hz, 10–15 mW) often elicited prototypical orofacial rejection behaviour (Supplementary Video 1).



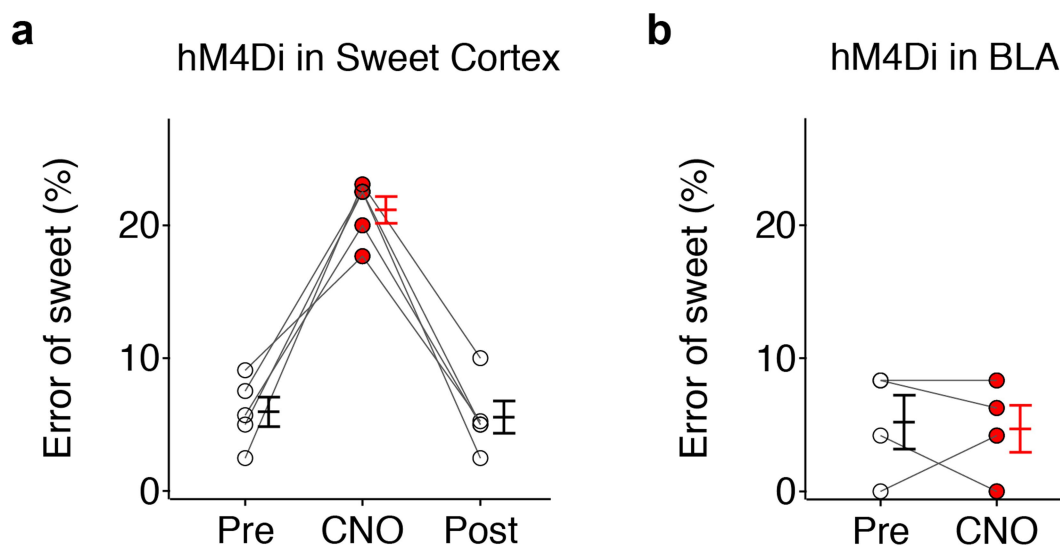
**Extended Data Fig. 4 | Activation of bitter cortical projections to posterior BLA is aversive.** As shown in Fig. 1 and Extended Data Fig. 1, a fraction of the bitter cortico-amygdalar projections terminate in the posterior BLA. As expected, stimulation of these projections elicits aversive responses. **a**, Representative histograms showing licking events in the presence (blue) or absence (grey) of photostimulation of bitter cortical

projections to the posterior BLA. AAV-ChR2 was injected into the bitter cortex, and the stimulating fibre was targeted above the posterior BLA (coordinates: bregma  $-2$  mm; lateral  $3.4$  mm; ventral  $4.3$  mm). **b**, Quantification of licking responses.  $n = 3$  mice, two-tailed paired  $t$ -test,  $P = 0.0121$ . Values are mean  $\pm$  s.e.m.



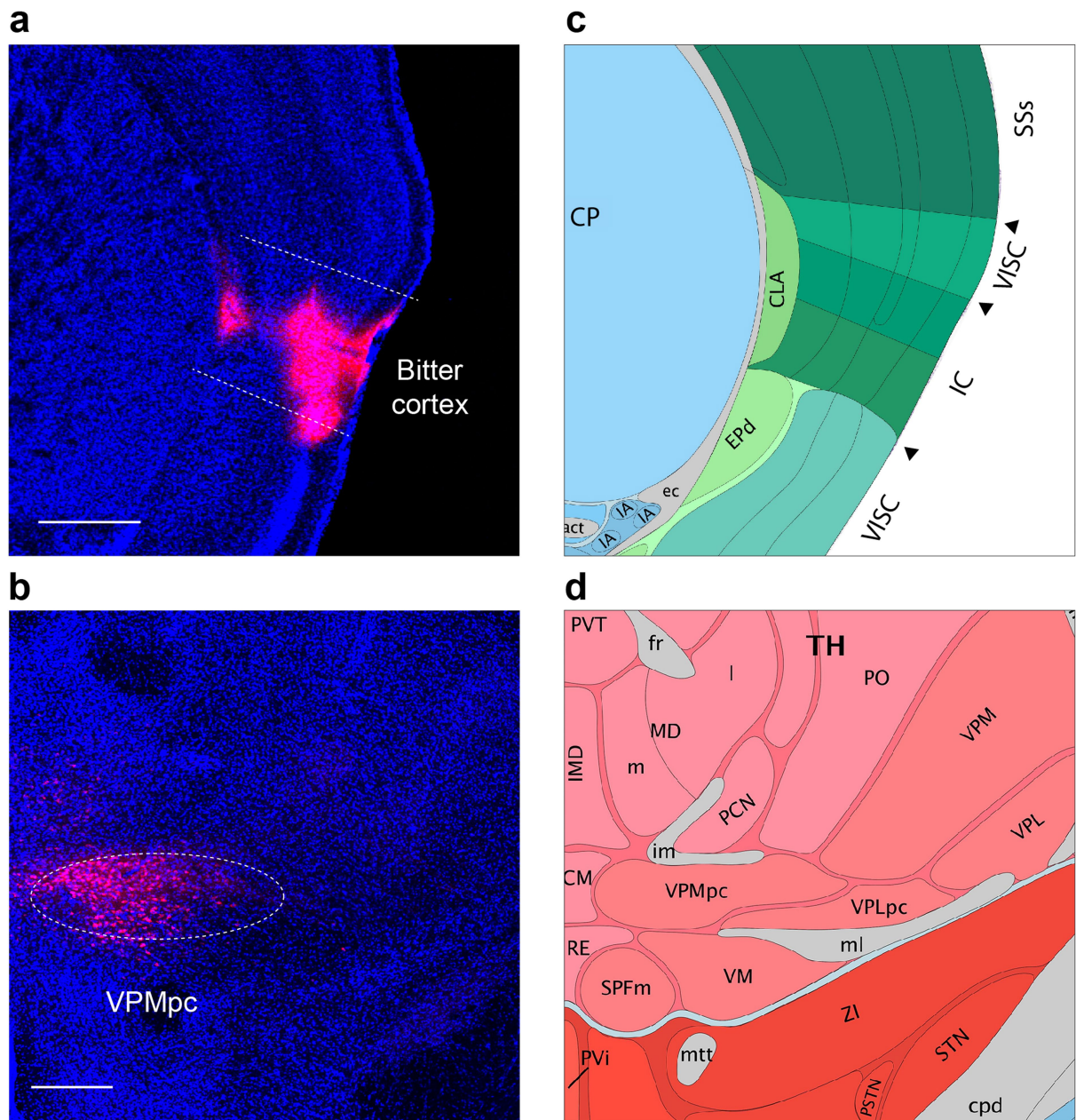
**Extended Data Fig. 5 | Control experiments for silencing amygdala with DREADD and NBQX. a,** Quantification of licking response before and after saline administration in mice that expressed inhibitory DREADD (hM4Di; see Fig. 5).  $n = 6$  mice, two-tailed paired  $t$ -test; pre,  $P = 0.0011$ ; saline,  $P < 0.0001$ ; post,  $P = 0.0014$ . **b,** Quantification of licking response before and after CNO administration ( $10 \text{ mg kg}^{-1}$ ) in wild-type (WT)

non-DREADD-expressing animals.  $n = 6$  mice, two-tailed paired  $t$ -test; pre,  $P = 0.0025$ ; CNO,  $P = 0.0008$ ; post,  $P = 0.0021$ . **c,** Controls with saline infusion instead of NBQX for Fig. 5c.  $n = 6$  mice, two-tailed paired  $t$ -test; pre,  $P = 0.0080$ ; saline,  $P = 0.0054$ ; post,  $P = 0.0046$ . Values are mean  $\pm$  s.e.m.



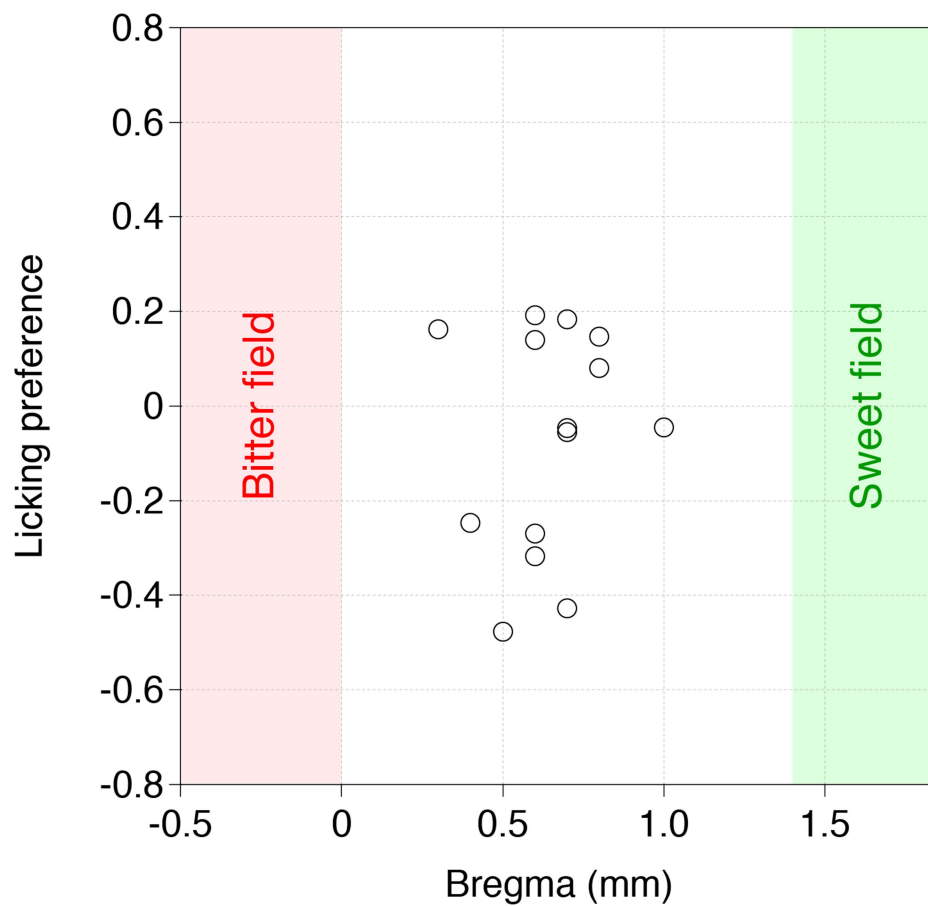
**Extended Data Fig. 6 | Chemogenetic silencing of neurons in the taste cortex impairs tastant recognition.** **a**, Quantification of the error rate for sweet taste recognition (2 mM AceK) in a three-port assay before and after silencing neurons in the sweet taste cortex with inhibitory DREADD (hM4Di) and CNO (10 mg kg<sup>-1</sup>).  $n = 5$  mice, repeated-measures one-way ANOVA followed by Bonferroni post hoc test,  $F_{2,8} = 46.84$ ,  $P < 0.0001$ . Pharmacological silencing data using NBQX can be found in a previously published study<sup>7</sup>. **b**, Quantification of the error rate of sweet taste

recognition (1 mM AceK) before and after silencing of the amygdala with inhibitory DREADD (hM4Di) and CNO (10 mg kg<sup>-1</sup>).  $n = 4$  mice, two-tailed paired  $t$ -test; pre versus CNO,  $P = 0.7888$ . See also Fig. 6. Note that sweet recognition was only affected by silencing the taste cortex, but not the amygdala. All tested animals recognized sweet taste with the correct behavioural choice (before silencing) in at least 90% of the trials. Values are mean  $\pm$  s.e.m. See Supplementary Table 1 for detailed statistics.



**Extended Data Fig. 7 | Retrograde labelling of gustatory thalamic neurons.** **a, b**, Injection of the retrograde tracer cholera toxin subunit B–Alexa Fluor 594 in the taste cortex (bitter cortical field; shown in red in **a**)

selectively labels neurons in the taste thalamus (VPMpc; **b**). Similar results were observed in two animals. **c, d**, Diagrams of the corresponding brain regions, adapted from the Allen Brain Institute atlas. Scale bars, 500  $\mu\text{m}$ .



**Extended Data Fig. 8 | Licking responses to photostimulation of intermediate regions between sweet and bitter cortex.** Behavioural responses (see Fig. 3) in water-only trials linked to contact-driven self-stimulation are shown for mice expressing ChR2 between the sweet and

bitter cortical fields. Note that a positive index means attraction, whereas a negative index means aversion to light stimulation.  $n = 14$  mice. Data points indicate the behavioural test of individual animals at different stimulation sites relative to bregma position.

# Self-organization of a human organizer by combined Wnt and Nodal signalling

I. Martyn<sup>1,2,3</sup>, T. Y. Kanno<sup>1,3</sup>, A. Ruzo<sup>1</sup>, E. D. Siggia<sup>2\*</sup> & A. H. Brivanlou<sup>1\*</sup>

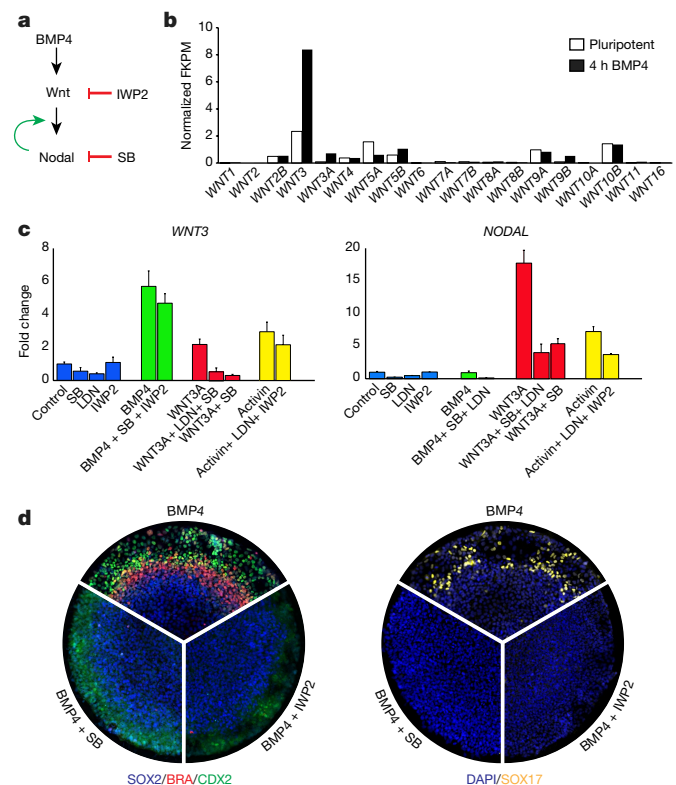
In amniotes, the development of the primitive streak and its accompanying ‘organizer’ define the first stages of gastrulation. Although these structures have been characterized in detail in model organisms, the human primitive streak and organizer remain a mystery. When stimulated with BMP4, micropatterned colonies of human embryonic stem cells self-organize to generate early embryonic germ layers<sup>1</sup>. Here we show that, in the same type of colonies, Wnt signalling is sufficient to induce a primitive streak, and stimulation with Wnt and Activin is sufficient to induce an organizer, as characterized by embryo-like sharp boundary formation, markers of epithelial-to-mesenchymal transition and expression of the organizer-specific transcription factor GSC. Moreover, when grafted into chick embryos, human stem cell colonies treated with Wnt and Activin induce and contribute autonomously to a secondary axis while inducing a neural fate in the host. This fulfils the most stringent functional criteria for an organizer, and its discovery represents a milestone in human embryology.

The pioneering experiments of Spemann and Mangold demonstrated that a small group of cells located on the dorsal side of the early amphibian embryo have the ability to induce and ‘organize’ a complete secondary axis when transplanted to the ventral side of another embryo<sup>2</sup>. This led to the concept of the ‘organizer’, and later discovery of embryonic tissue with similar organizer activity in fish, birds, and rodents<sup>3–6</sup> demonstrated that this early embryonic activity is evolutionarily conserved. Cells in the organizers of all species studied to date exhibit the same behaviour during axis induction: they (i) contribute autonomously to axial and paraxial mesoderm, including head process and notochord, and (ii) induce neural fate non-autonomously in their neighbours. The organizer in primates, including humans, has so far not been defined.

Owing to the ethical limitations of working with early human embryos, the only way to search for the human organizer is via human embryonic stem cells (hESCs). We have previously shown that when grown on geometrically confined discs, hESCs respond to BMP4 by differentiating and self-organizing into concentric rings of embryonic germ layers: with ectoderm in the centre, extra-embryonic tissue at the edge, and mesoderm and endoderm in between<sup>1</sup>. These embryo-like ‘gastruloids’ are robust and amenable to analysis with subcellular resolution.

During mouse gastrulation, BMP4 signalling activates the Wnt pathway which in turn activates the Activin–Nodal pathway (Fig. 1a), at both the transcriptional and signalling levels<sup>7</sup>. Since it has been shown in mouse and other vertebrates that these three pathways are the most critical pathways for organizer formation<sup>8,9</sup>, we first investigated whether this hierarchy was conserved in human gastruloids. Using RNA sequencing analysis (RNA-seq), we found that among the 19 Wnt genes in the human genome, only *WNT3* is markedly and immediately induced upon stimulation with BMP4 (Fig. 1b). Reverse transcription with quantitative PCR (qPCR) analysis shows that activation of Wnt signalling directly induces *NODAL* expression (Fig. 1c). Further qPCR

analysis showed that *NODAL* induction was reduced when the Nodal inhibitor SB431542 (SB) was present. Together with the observation that Activin induces *NODAL* expression, this suggests the presence of



**Fig. 1 | Primitive streak signalling in hESCs follows the BMP4 to Wnt to Nodal hierarchy.** **a**, Model of the proposed hierarchy of signalling that initiates the primitive streak in hESCs, along with indication of the steps at which the inhibitors SB431542 (SB) and IWP2 may act. As in the mouse, BMP4 acts on Wnt, and Wnt then acts on Nodal. There is also positive feedback between Wnt and Nodal. **b**, RNA-seq analysis of Wnt ligand expression, in pluripotency and after 4 h of stimulation with BMP4, in 500- $\mu$ m diameter hESC micropatterns. The results show that overall Wnt expression is low in the pluripotent state and that *WNT3* is the only strong and direct Wnt target of BMP4 stimulation. Data are from a previously published dataset<sup>19</sup> (GEO accession number GSE77057). FKPM, Fragments per kilobase million. **c**, qPCR analysis showing expression of *WNT3* and *NODAL* in micropatterned hESC colonies after 4 h of stimulation as indicated. Data are mean  $\pm$  s.d. of  $n = 3$  biologically independent replicates. LDN, LDN193189. **d**, Pie sectors are of representative 1,000- $\mu$ m-diameter micropatterned hESC colonies stimulated with BMP4, BMP4 + IWP2, or BMP4 + SB. Colonies were fixed 48 h after stimulation and stained for germ layer molecular markers. All micropattern experiments were performed on at least three separate occasions with similar results, and unless mentioned otherwise, all other micropatterns shown are 1,000  $\mu$ m in diameter. Staining is quantified in Extended Data Fig. 1c.

<sup>1</sup>Laboratory of Molecular Vertebrate Embryology, The Rockefeller University, New York, NY, USA. <sup>2</sup>Center for Studies in Physics and Biology, The Rockefeller University, New York, NY, USA. <sup>3</sup>These authors contributed equally: I. Martyn, T. Y. Kanno. \*e-mail: siggiae@rockefeller.edu; brvnlou@rockefeller.edu

a Nodal feedback loop, as in the mouse<sup>10</sup>. Additionally, no direct BMP4 induction by either Wnt or Nodal signalling was observed (Extended Data Fig. 1b). This demonstrates that the transcriptional hierarchy of BMP4 to WNT3 to Nodal is evolutionarily conserved in hESCs.

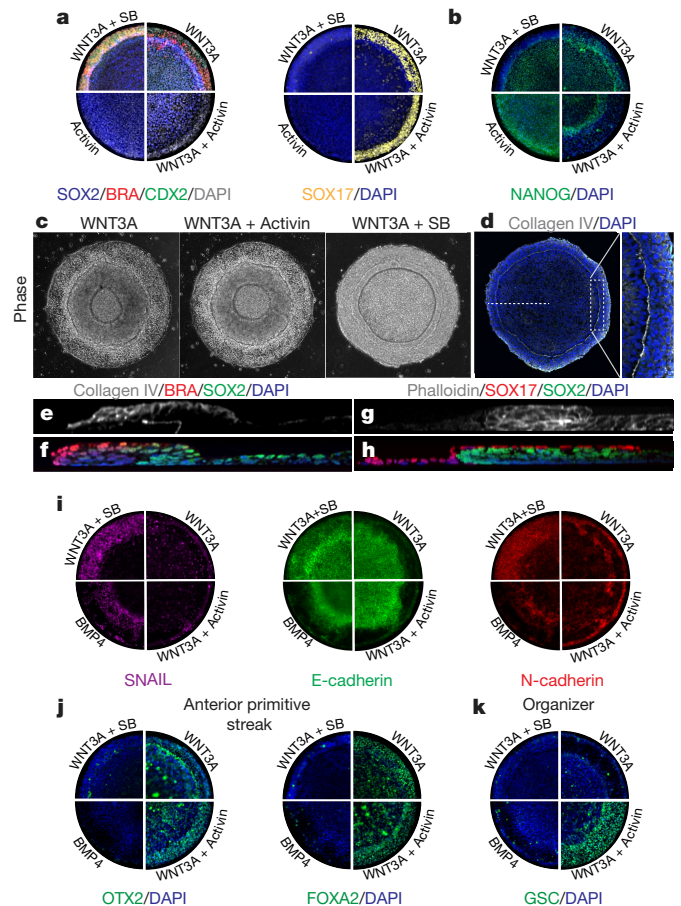
To establish whether the hierarchy of signalling activity was also conserved, we challenged the self-organizing activity of BMP4 with SB and the Wnt inhibitor IWP2. Neither inhibitor had an effect in the absence of BMP4, but either IWP2 or SB prevented BMP4-induced formation of mesoderm (indicated by BRA expression) and endoderm (indicated by SOX17 expression; Fig. 1d and Extended Data Fig. 1c). Thus both Wnt and Activin–Nodal signalling are necessary for mesendodermal induction and patterning downstream of BMP4.

To determine whether either Wnt or Activin–Nodal signalling alone were sufficient as well as necessary for the organizing activity of BMP4, hESC colonies were stimulated with either WNT3A or Activin. After 48 h of treatment with WNT3A, cells at the periphery of colonies differentiated into mesoderm and endoderm (Fig. 2a, Extended Data Fig. 2a), whereas cells in the centre of colonies maintained their pluripotent epiblast fate (indicated by SOX2 and NANOG expression) (Fig. 2a, b) rather than differentiating into ectoderm. After 48 h of treatment with Activin, however, the cells showed no sign of differentiation or self-organization, and all maintained the same morphology and expression of pluripotency markers (Fig. 2a, b).

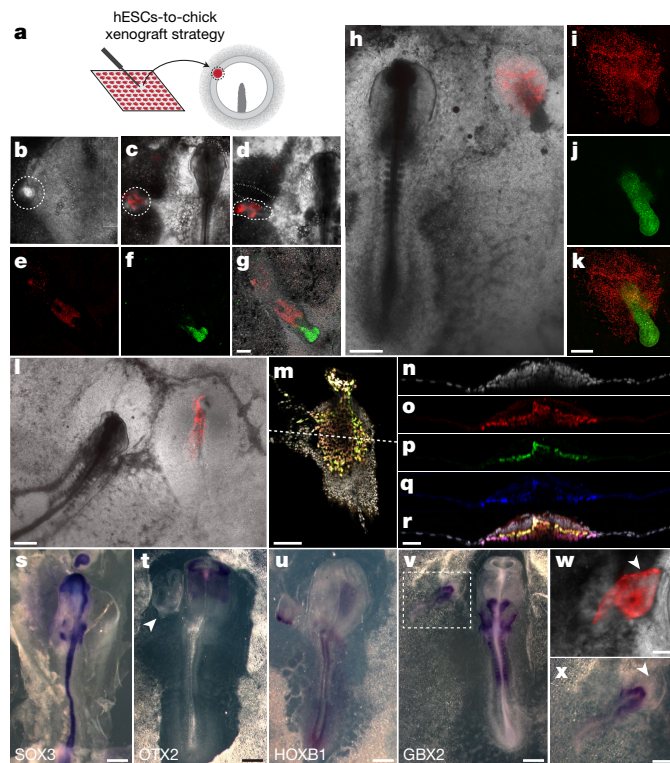
As it is unlikely that Activin–Nodal has no effect during human gastrulation, we presented WNT3A in two combinations that represent the opposite extremes of an Activin–Nodal gradient: either as WNT3A + Activin or as WNT3A + SB. Consistent with studies in model systems and human and mouse embryonic stem cells<sup>11,12</sup>, we found that Activin–Nodal signalling acts as a modifier of mesoderm and endoderm patterning—in the presence of Activin, cells on the periphery converted to endoderm (SOX17+) with no mesoderm (BRA–) expression, whereas in the presence of SB all cells converted to mesoderm (BRA+), with no endoderm (SOX17–) expression (Fig. 2a; Extended Data Fig. 2a). As with WNT3A treatment alone, both sets of colonies had sharp morphological boundaries that became much more pronounced after 48 h (Fig. 2c). The expression profile of Collagen IV<sup>13</sup> and the 3D organization of cells around this boundary (Fig. 2d–h) was highly reminiscent of the morphological signature of a primitive streak. Confirming the primitive streak-like nature of these structures, we found evidence of an epithelial-to-mesenchymal transition (EMT), with expression of SNAIL and a switch from expression of E-cadherin to N-cadherin, in which the mesendodermal fates are later established (Fig. 2i, Extended Data Fig. 2b). Taken together, these results show that Wnt signalling is necessary and sufficient to induce a primitive streak, and that Activin–Nodal signalling acts as a modifier that controls the timing of EMT and patterning of mesoderm versus endoderm.

Having identified primitive streak in our human gastruloids, we next investigated whether an organizer subpopulation was also present. In mouse, the organizer is located in the anterior primitive streak in what is thought to be the region with highest Nodal signalling<sup>14</sup>. We found that treatment with either WNT3A alone or with combined WNT3A and Activin results in co-expression of OTX2 and FOXA2 in the same micropattern region that expresses SOX17. This combination of markers is characteristic of anterior primitive streak in mouse. However, as in the mouse, only the condition with the highest Nodal signalling, that is, combined treatment with WNT3A and Activin, results in the expression of the organizer-specific marker GSC (Fig. 2j, k, Extended Data Fig. 3c). This combined treatment also leads to the highest expression of genes for key secreted inhibitors that are known to be produced by the organizer and its derivatives<sup>15</sup>, such as *CHRD*, *DDK1*, *CER1*, *LEFTY1* and *LEFTY2*, as well as the highest expression of *NODAL*, which at later stages in mouse is also specific to the organizer (Extended Data Fig. 4d).

The induction of characteristic organizer markers, the emergence of a sharp Collagen IV-based morphological boundary dividing the primitive streak and epiblast regions, and the induction of EMT are all evidence that support induction of a human organizer in an early primitive



**Fig. 2 | Wnt is necessary and sufficient to induce primitive streak markers and morphology.** **a**, Micropatterned hESC colonies were stimulated with WNT3A, WNT3A + Activin, WNT3A + SB, or Activin, fixed 48 h after stimulation and stained for germ layer molecular markers. Staining is quantified in Extended Data Fig. 2a. **b**, Micropatterned hESC colonies were stimulated with WNT3A, WNT3A + Activin, WNT3A + SB, or Activin, and fixed 48 h after stimulation and stained for the pluripotency marker NANOG. **c**, Sharp boundaries are visible even by phase contrast microscopy 52 h after colonies were stimulated with WNT3A, WNT3A + SB, or WNT3A + Activin. **d–f**, Micropatterned hESC colonies stained with DAPI (blue; **d** and **f**), Collagen IV (white; **d** and **e**), BRA (red; **f**), and/or SOX2 (green; **f**) 52 h after stimulation with WNT3A + SB. Note how Collagen IV traces a sharp continuous circle line between the periphery and the interior of the colony and forms a dividing line between the epiblast-like SOX2+ interior and the differentiating BRA+ primitive streak exterior. In addition, our EMT data shows that the BRA– and SOX2– cells express SNAIL. **g, h**, Micropatterned hESC colonies stained with phalloidin (white; **g**) or DAPI (blue; **h**), SOX17 (red; **h**) or SOX2 (green; **h**) 52 h after stimulation with WNT3A. The cross section shows an inner epithelial SOX2+ epiblast region folded back in on itself. Outside and above the fold is the mesenchymal SOX17 region. This geometry seems to permit the mesenchymal SOX17 cells to migrate along the epiblast-like region in basal-to-basal contact. Consistent with this, no SOX17 cells were observed further inwards beyond the leading edge of the fold where the basal presenting surface ends. **i**, Micropatterned hESC colonies were stimulated with BMP4, WNT3A, WNT3A + SB, or WNT3A + Activin, fixed 48 h after stimulation and stained for the EMT markers SNAIL, E-cadherin and N-cadherin. Colonies stimulated with WNT3A, and those stimulated with WNT3A + Activin exhibited lower SNAIL expression after 48 h because they entered EMT earlier, at 24 h (see Extended Data Fig. 2b for 12, 24 and 36 h time points). **j, k**, Micropatterned hESC colonies were stimulated with WNT3A, WNT3A + Activin, WNT3A + SB, or BMP4, fixed 48 h after stimulation and stained for markers characteristic of anterior primitive streak (OTX2 and FOXA2; **j**) or the organizer-specific transcription factor GSC (**k**). Staining is quantified in Extended Data Fig. 3a. All experiments were repeated at least three times independently with similar results.



**Fig. 3 | Human organizer induces secondary axis in chick embryo.** **a**, Schematic showing the strategy for the hESC-chick xenograft. **b–g**, Secondary axis induced by a micropatterned RUES2-GLR colony, stimulated for 24 h, and grafted into a HH3-stage chick. SOX17-tdTomato (red) live marker at 0, 24 and 38 h after grafting (**b–d**); SOX2 (green) is prominent in the tip of the secondary axis 48 h after grafting (**f, g**), and does not overlap with the hESCs (red, human nuclear antigen (HNA); **e, g**). Scale bar, 200  $\mu$ m. **h–k**, A different 24-h-stimulated micropatterned hESC colony inducing a secondary axis in a chick host 27 h after grafting. Image of SOX17-tdTomato hESC cells in the live grafted embryo (red, **h**). The fixed embryo, stained for HNA (red; **i, k**) and SOX2 (green; **j, k**). Scale bars, 500  $\mu$ m (**h**), 200  $\mu$ m (**i–k**). **l–r**, Example of secondary axis induction from a 24-h-stimulated micropatterned hESC colony with more complete self-organizing structures, 27 h after grafting. **l**, Image of SOX17-tdTomato hESC cells in the live grafted embryo (red). **m**, Confocal section of secondary axis stained with DAPI (grey), HNA (red) and BRA (green). **n–r**, Confocal projection of the region indicated by the line in **m**, also showing staining for SOX17 (blue). The merged image (**r**) shows how the secondary axis is layered, with epiblast chick cells on top of a layer of human BRA+ cells, which in turn rest on a layer of human SOX17+ cells, in a similar arrangement to that of epiblast, mesoderm and endoderm layers in a gastrulating mouse or chick embryo. Scale bars, 500  $\mu$ m (**l**), 100  $\mu$ m (**m**) and 50  $\mu$ m (**n–r**). **s–v**, RNA in situ hybridization in the grafted embryo. **s**, Chicken SOX3 is expressed throughout the neural tube and head in the host chick, as well as in the induced secondary axis. **t**, OTX2 is expressed in the host forebrain but is absent in the graft-induced tissue (indicated by the arrowhead). **u**, HOXB1 is expressed in the host and in the graft-induced secondary axis. **v**, GBX2 is expressed in the host and in the graft-induced secondary axis. **w, x**, Magnified view of the region indicated in **v**; GBX2 mRNA expression (**x**), and the secondary axis and tdTomato-hESCs (red) in the fixed tissue (**w**). The arrow shows the location of the hESC graft. Scale bars, 500  $\mu$ m (**s–v**), 250  $\mu$ m (**w, x**). All experiments were performed at least three times with similar results; exact numbers of replicates and measures of reproducibility are shown in Table 1.

streak by treatment with WNT3A and Activin. However, as originally defined by the classic amphibian experiments, an organizer is determined functionally as a group of cells that can induce a secondary axis when grafted ectopically into host embryos<sup>2</sup>. In this context, the grafted cells should contribute directly to the ectopic axis (autonomously),

**Table 1 | Induction of chick neural tissue by hESC micropatterns**

	Survival	SOX2	SOX3	GBX2	HOXB1	OTX2
Control (medium only)	4/15	0/15	—	—	—	—
WNT3A + Activin, 24 h	273/300****	10/19****	—	5/6	9/10	0/14
WNT3A + Activin, 48 h	37/40****	6/15****	6/8	—	—	—
WNT3A, 48 h	8/15	0/15	—	—	—	—
WNT3A + SB, 48 h	7/15	0/15	—	—	—	—
BMP4, 48 h	14/15***	0/15	—	—	—	—

Survival column indicates whether treated and grafted hESC cells from micropatterned RUES2-GLR colonies were detected in the live host chick 24 h after the graft. Many of these grafts were used to optimize the antibodies and probes listed in the remaining columns. Example control grafts are shown in Extended Data Fig. 7. Statistical analysis ( $\chi^2$ ,  $2 \times 2$  contingency test, compared to the control condition). \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ , \*\*\*\* $P < 0.0001$ .

and induce neural tissue in the cells of the host (non-autonomously). In order to test for the most stringent and functional definition of an organizer, we used an ex ovo cross-species transplantation strategy based on previous mammalian organizer studies<sup>16,17</sup>, grafting fluorescent reporter micropatterned hESC colonies treated for 24 or 48 h with WNT3A and Activin into the marginal zone of early chick culture embryos<sup>18</sup> (stage HH2 to HH3+). We used 500- $\mu$ m diameter micropatterns rather than 1,000- $\mu$ m diameter micropatterns as these had a higher proportion of GSC+ cells, and we performed grafts at both 24 h and 48 h after treatment, as GSC first becomes apparent and is co-expressed with BRA 24 h after treatment (Extended Data Fig. 4a–c). For the reporter line, we used the CRISPR-Cas9 generated RUES2-GLR (germ layer reporter) cell line (see Extended Data Figs. 5, 6 and Methods).

We found that RUES2-GLR grafts survived, mingled with host cells, and induced and contributed to a secondary axis that became apparent between 24 and 48 h after grafting (Fig. 3b–l). Both the live cell reporter and a human-specific nuclear antigen (HNA) revealed that the human cells directly contributed to the ectopic axis autonomously and continued to differentiate in their new environment, contributing both BRA+ and SOX17+ cells (Fig. 3h, l, m). This mirrors previous observations in mouse-to-mouse organizer grafting experiments<sup>5</sup>. Confocal cross-sections of these secondary axes revealed self-organizing features directly resembling those found in the early chick and mouse embryo, including correct layering of germ layers and central elongated notochord-like structures composed partly or entirely of graft-derived cells (Fig. 3n–r, Extended Data Fig. 8, and Supplementary Video 1). Analysis of molecular markers also established that the human cells induced neural tissue in the chick non-autonomously: SOX2 and SOX3 were ectopically induced in chick cells that surrounded the human cells (Fig. 3e–g, i–k, s). RNA in situ hybridization and antibody staining for HOXB1, GBX2 and OTX2 established that the neural tissue was predominantly posterior in nature (Fig. 3t–v). Since in the mouse the early-gastrula organizer and late-streak node also do not induce anterior neural structures when grafted to another mouse embryo, this result suggests that our human organizer is closer to these organizer stages than to the mouse mid-gastrula organizer<sup>6,15</sup>. As controls, RUES2-GLR grafts derived from hESCs treated with WNT3A, WNT3A plus SB, BMP4, or medium alone exhibited lower survival rates and did not induce chick neural markers (Table 1 and Extended Data Fig. 5a). Taken together with the morphological, cellular and molecular evidence described above, this functional test in an embryonic environment provides the most stringent evidence for the induction of a human organizer. It also highlights that the organizer itself can be obtained in vitro by self-organization of hESCs in response to WNT3A and Activin treatment in a confined micropattern geometry.

Our ability to generate a human organizer closes the loop that was initiated by classical experimental embryologists working on amphibian systems nearly 100 years ago, and demonstrates that the concept of the organizer is evolutionarily conserved from frogs to humans. Our chick experiments also define an in vivo platform to

validate results obtained in an in vitro gastruloid platform, and may be generally applicable to test and explore other aspects of early human development.

### Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0150-y>.

Received: 25 April 2017; Accepted: 16 April 2018;

Published online: 23 May 2018

1. Warmflash, A., Sorre, B., Etoc, F., Siggia, E. D. & Brivanlou, A. H. A method to recapitulate early embryonic spatial patterning in human embryonic stem cells. *Nat. Methods* **11**, 847–854 (2014).
2. Spemann, H. & Mangold, H. Induction of embryonic primordia by implantation of organizers from a different species. *Int. J. Dev. Biol.* **45**, 13–38 (2001).
3. Oppenheimer, J. M. Transplantation experiments on developing teleosts (*Fundulus* and *Perca*). *J. Exp. Zool.* **72**, 409–437 (1936).
4. Waddington, C. H. Experiments on the development of chick and duck embryos, cultivated in vitro. *Philos. Trans. R. Soc. London. B* **221**, 179–230 (1932).
5. Beddington, R. S. Induction of a second neural axis by the mouse node. *Development* **120**, 613–620 (1994).
6. Kinder, S. J. et al. The organizer of the mouse gastrula is composed of a dynamic population of progenitor cells for the axial mesoderm. *Development* **128**, 3623–3634 (2001).
7. Ben-Haim, N. et al. The Nodal precursor acting via Activin receptors induces mesoderm by maintaining a source of its convertases and BMP4. *Dev. Cell* **11**, 313–323 (2006).
8. Crease, D. J., Dyson, S. & Gurdon, J. B. Cooperation between the Activin and Wnt pathways in the spatial control of organizer gene expression. *Proc. Natl Acad. Sci. USA* **95**, 4398–4403 (1998).
9. Gritsman, K., Talbot, W. S. & Schier, A. F. Nodal signaling patterns the organizer. *Development* **127**, 921–932 (2000).
10. Brennan, J. et al. Nodal signalling in the epiblast patterns the early mouse embryo. *Nature* **411**, 965–969 (2001).
11. Zorn, A. M. & Wells, J. M. Vertebrate endoderm development and organ formation. *Annu. Rev. Cell Dev. Biol.* **25**, 221–251 (2009).
12. Faial, T. et al. Brachyury and SMAD signalling collaboratively orchestrate distinct mesoderm and endoderm gene regulatory networks in differentiating human embryonic stem cells. *Development* **142**, 2121–2135 (2015).

13. Williams, M., Burdsal, C., Periasamy, A., Lewandoski, M. & Sutherland, A. Mouse primitive streak forms in situ by initiation of epithelial to mesenchymal transition without migration of a cell population. *Dev. Dyn.* **241**, 270–283 (2012).
14. Tam, P. P. L. & Loebe, D. A. F. Gene function in mouse embryogenesis: get set for gastrulation. *Nat. Rev. Genet.* **8**, 368–381 (2007).
15. Robb, L. & Tam, P. P. L. Gastrula organiser and embryonic patterning in the mouse. *Semin. Cell Dev. Biol.* **15**, 543–554 (2004).
16. Zhu, L., Belo, J. A., De Robertis, E. M. & Stern, C. D. Goosecoid regulates the neural inducing strength of the mouse node. *Dev. Biol.* **216**, 276–281 (1999).
17. Knoetgen, H., Teichmann, U., Wittler, L., Viebahn, C. & Kessel, M. Anterior neural induction by nodes from rabbits and mice. *Dev. Biol.* **225**, 370–380 (2000).
18. Chapman, S. C., Collignon, J., Schoenwolf, G. C. & Lumsden, A. Improved method for chick whole-embryo culture using a filter paper carrier. *Dev. Dyn.* **220**, 284–289 (2001).
19. Etoc, F. et al. A balance between secreted inhibitors and edge sensing controls gastruloid self-organization. *Dev. Cell* **39**, 302–315 (2016).

**Acknowledgements** The authors are grateful to I. Yan, F. Vieceli and M. Bronner for materials and protocols, to J. Metzger for assistance with 3D image segmentation, and to members of the A.H.B. and E.D.S. laboratories for helpful discussions. This work was supported by grants R01 HD080699, R01 GM101653, the Tri-Institutional Starr Foundation Grant 2016-007, and private funds from the Rockefeller University.

**Reviewer information** Nature thanks I. Hyun, O. Pourqu   and the other anonymous reviewer(s) for their contribution to the peer review of this work.

**Author contributions** I.M., A.H.B. and E.D.S. conceptualized the work and wrote the paper. I.M. performed stem cell experiments. I.M. devised, and I.M. and T.Y.K. performed chick experiments. A. R. conceived, generated and validated the RUES2-GLR cell line. All authors reviewed the manuscript.

**Competing interests** The authors declare no competing interests.

### Additional information

**Extended data** is available for this paper at <https://doi.org/10.1038/s41586-018-0150-y>.

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41586-018-0150-y>.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>.

**Correspondence and requests for materials** should be addressed to E.D.S. or A.H.B.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized, and the investigators were not blinded to allocation during experiments and outcome assessment.

**Ethics statement.** This work was conducted according to protocols approved by the Tri-Institutional Stem Cell Initiative Embryonic Stem Cell Research Oversight Committee (Tri-SCI ESCRO), an independent committee charged with oversight of research with human pluripotent stem cells and embryos to ensure conformance with university policies, and guidelines from the US National Academy of Sciences (NAS) and the International Society for Stem Cell Research (ISSCR). The Tri-SCI ESCRO is composed of members with scientific and bioethical expertise. The ESCRO review of these protocols was conducted before the May 2016 issuance of the ISSCR guidelines, but the review addressed the issues of growth and culture of human–chick chimaeras and in vitro culture of embryo-like structures and anticipated the ISSCR guidelines (specifically recommendations 2.1.3 and 2.1.5, which are pertinent to this study). As part of these protocols, the human cells transplants were limited to <10% compared to host animal at any given stage, and no chicken–human chimaeras were allowed to hatch. Additionally, the researchers considered that the self-organized structures that arose from the experiments lacked human organismal potential owing to the in vitro culture lacking the necessary non-embryonic tissues or support that are present in vivo. The ESCRO Committee also reviewed and approved the NIH grants HD080699 and GM101653 that funded part of this study, and approved the initial derivation of the RUES2 cell line which is listed in the NIH Human Embryonic Stem Cell Registry.

**Cell culture.** hESCs (RUES2 cell line) were grown in HUESM medium conditioned by mouse embryonic fibroblasts (MEF-CM) and supplemented with 20 ng/ml bFGF. Mycoplasma testing was carried out before beginning experiments and again at two-monthly intervals. For maintenance, cells were grown on GelTrex (1:40 dilution, Invitrogen)-coated tissue culture dishes (BD Biosciences). The dishes were coated overnight at 4 °C and then incubated at 37 °C for at least 20 min before the cells were seeded. Cells were passaged using Gentle Cell Dissociation Reagent (Stem Cell Technologies, 07174).

**Micropatterned cell culture.** Micropatterned coverslips were made according to a new protocol devised in our laboratory that reduced operating costs. First, 22 × 22 mm no. 1 coverslips were spin-coated with a thin layer of polydimethylsiloxane (PDMS; Momentive, RTV615A) and left to set overnight. They were then coated with 5 µg/ml laminin 521 (Biolamina) diluted in PBS with calcium and magnesium (PBS++) for 2 h at 37 °C. After two washes with PBS++, coverslips were placed under a positive feature UV Quartz Mask (Applied Image) in a home-made UV oven. Laminin not protected by the features in the mask was burned off by 10 min of deep UV application (185-nm wavelength). Coverslips were then removed, washed twice more with PBS++, and then left at 4 °C overnight in 1% F127-Pluronic (Sigma) solution in PBS++. The now patterned coverslips were used within one week of fabrication. To seed cells onto a micropatterned coverslip, cells were dissociated from growth plates with StemPro Accutase (Life Technologies) for 7 min. Cells were washed once with growth medium, washed again with PBS, and then re-suspended in growth medium with 10 µM ROCK inhibitor Y-27632 (Abcam). Coverslips were placed in 35-mm plastic tissue culture dishes, and 1 × 10<sup>6</sup> cells in 2 ml of medium were used for each coverslip. After 1 h the ROCK inhibitor was removed and replaced with standard growth medium supplemented with penicillin–streptomycin (Life Technologies). Cells were stimulated or treated with the following ligands or small molecules 12 h after seeding: 100 ng/ml WNT3A, 50 ng/ml BMP4, 100 ng/ml Activin, 10 µM SB431542, 2 µM IWP2 or 200 nM LDN193189.

**hESC immunocytochemistry.** Cells were fixed in 4% paraformaldehyde for 20 min at room temperature, washed twice with PBS, then blocked and permeabilized with 3% donkey serum and 0.1% Triton X-100 in PBS for 30 min, also at room temperature. Cells were incubated overnight with primary antibodies in this blocking buffer at 4 °C. The next day they were washed three times with PBS + 0.1% Tween-20 for 30 min each, then incubated with secondary donkey antibodies (Alexa Fluor 488, Alexa Fluor 555, Alexa Fluor 647) and DAPI for 30 min. Cells were then washed twice more with PBS, and mounted on glass slides for imaging. Primary antibodies used are listed in Supplementary Table 1.

**Chick immunofluorescence.** Embryos were fixed in 4% paraformaldehyde in PBS for 1 h at room temperature or overnight at 4 °C. They were then washed three times with PBST (PBS + 0.5% Triton X-100) for 1 h each on a nutator and blocked and permeabilized with 3% donkey serum, 1% bovine serum albumin in PBST for 2 h, also at room temperature. Next, they were incubated overnight at 4 °C with anti-SOX2 antibody (R&D, AF2018) diluted in blocking buffer. The next day embryos were washed three times with PBST for 1 h each on a nutator and then incubated with secondary donkey antibody Alexa Fluor 594, anti-human nuclear antigen (HNA) (Novus Biologicals, NBP2-34525AF647), and DAPI overnight. Embryos were washed three times with PBST for 1 h each and mounted on glass slides with fluoromount for imaging.

**Chick RNA in situ hybridization.** Chicken SOX3 probe was kindly provided by F. M. Vieceli, and whole-mount RNA in situ hybridization was performed using previously described procedures<sup>20</sup>. In brief, the embryos were fixed overnight in 4% paraformaldehyde in PBS 24–48 h after the grafting. The embryos were then washed three times with PBS + 0.1% Tween-20, and then dehydrated through a methanol series (25% methanol/PBS, 50% methanol/PBS, 75% methanol/PBS, 100% methanol) and rehydrated (100% methanol, 75% methanol/PBS, 50% methanol/PBS, 25% methanol/PBS), in 15-min steps at room temperature. Next, the embryos were incubated with 10 µg/ml Proteinase K for 5 min, rinsed twice in PBS + 0.1% Tween-20, incubated in 2 mg/ml glycine in PBS + 0.1% Tween-20, washed two times in PBS + 0.1% Tween-20 for 5 min each and post-fixed for 20 min in 4% paraformaldehyde + 0.2% glutaraldehyde in PBS. The embryos were then hybridized at 70 °C using antisense RNA probes for chicken SOX3, OTX2, HOXB1 or GBX2 labelled with digoxigenin-11-UTP. The probe was localized using alkaline phosphatase-conjugated antibodies and the signal was developed with BM-purple.

**Microscopy and image analysis.** Images were acquired with either a Zeiss Axio Observer and a 20 × 0.8-NA lens, or with a Leica SP8 inverted confocal microscope with a 40 × 1.1-NA water-immersion objective. Image analysis and stitching was performed with ImageJ and custom Matlab routines. Images used in Supplementary Video 1 and Extended Data Fig. 8 were also deconvolved with Autoquant software and analysed in Imaris. In these images the notochord-like feature was identified by a combination of manual and Ilastick classification based on DAPI morphology, and cells belonging to this structure were segmented and false-coloured with the assistance of custom Python 3D segmentation software written by J. Metzger.

**qPCR data.** RNA was collected in Trizol at indicated time points from either micropatterned colonies or from small un-patterned colonies. Total RNA was purified using the RNeasy mini kit (Qiagen). qPCR was performed as described previously<sup>19</sup>. Primer sequences are listed in Supplementary Table 2.

**Transplantation of human organizer into chick host.** Fertilized white leghorn chicken eggs were incubated at 37–38 °C and 50% humidity and staged according to Hamburger and Hamilton<sup>21</sup>. Chick embryos were removed from the egg and set up in early-chick culture<sup>18</sup>, with Pannett–Compton saline solution as final wash and residual liquid in the culture. Instead of growing on home-made micropatterned coverslips, hESCs were grown on EMB CYTOO coverslips, as these had 500-µm diameter micropatterns. All other culture details were as described. Once grown to the indicated time with the indicated stimulation conditions, 500-µm diameter colonies were peeled off whole with tungsten needles (Fine Science Tools). These colonies were washed twice with Pannett–Compton solution to remove culture growth factors and ligands. Colonies were then moved to chick embryos and grafted into the marginal zone between the area opaca and area pellucida approximately 90° away from the site of primitive streak initiation, following the example of a typical Hensen's node graft<sup>22</sup>. The grafted embryos were then returned to the incubator to develop and were imaged live one day later and were ultimately fixed between 24 and 48 h after the graft. Owing to background from the agar mount and chick autofluorescence, only the SOX17–tdTomato marker had sufficient signal to noise ratio to be imaged live. In all steps penicillin–streptomycin was used to minimize the chances of bacterial contamination. I.M. and T.Y.K. were able to replicate the human–chick grafts independently, and the grafts could also be replicated by an experienced embryologist in our lab who was not involved in this study.

**Generation and validation of RUES2–GLR line.** CRISPR–Cas9 technology was used to generate a single hESC line containing three independent fate reporters (SOX2–mCitrine, BRA–mCerulean and SOX17–tdTomato). The already established and registered RUES2 hESC line (NIHhESC-09-0013) was used as the parental line. In order to achieve three independent targeting events in the same line, we approached each gene sequentially, since the efficiencies of recombination were not high enough for simultaneous targeting. First, for SOX17 targeting, we generated a homology donor plasmid (pSOX17–HomDon) containing: (i) a left homology arm containing a 1-kb sequence immediately upstream of the SOX17 stop codon; (ii) a P2A–H2B–tdTomato cassette; (iii) a floxed Neomycin selection cassette (loxP–PGK–Neo–pA–loxP); and (iv) a right homology arm containing a 1-kb sequence immediately downstream of the SOX17 stop codon. Note that, since the H2B–tdTomato is separated from the SOX17 gene by a self-cleaving P2A peptide, the expressed fluorescent reporter is not fused to SOX17, and therefore it will only be a reporter of the activation of the SOX17, as the two proteins may have different half-lives. We initially tried to use a direct fusion of SOX17 and tdTomato, but the fusion made the protein not localize correctly to the nucleus, and we therefore decided to use a self-cleaving strategy. All DNA fragments were amplified from pre-existing plasmids or genomic DNA using Q5 polymerase with the primers listed in Supplementary Table 3, and joined together using standard DNA ligation protocols. A single-guide RNA (sgRNA) recognizing a sequence near the stop codon of SOX17 (Supplementary Table 4) was cloned into a Cas9–nickase expression vector (pX335 from the Zhang laboratory, Addgene plasmid #42335).

This plasmid, together with the pSOX17-HomDon plasmid, was nucleofected into RUES2 cells using a Nucleofector II instrument and Cell Line Nucleofector Kit L (Lonza). Geneticin (a neomycin analogue) was added to the cultures five days after nucleofection, and maintained in the medium for a further seven days to ensure selection of correctly targeted clones. Colonies derived from single geneticin-resistant cells were picked and expanded for screening. PCR amplification and Sanger sequencing were used to identify correctly heterozygously targeted clones, with no unwanted mutations in the SOX17-sgRNA target site, both in the targeted and in the untargeted alleles. Positive clones were also validated for karyotyping (Giemsa banding). The top four potential off-target sites of the sgRNA were PCR amplified and Sanger sequenced to ensure no unwanted mutations were present. The pluripotency status and absence of differentiation of the clones were validated through immunofluorescence staining. Once a validated SOX17-tdTom clone was identified, its youngest frozen stock was thawed to undergo BRA gene targeting. Targeting of BRA followed a similar strategy as SOX17, but using a Puromycin resistance cassette. Colonies derived from single puromycin-resistant cells were screened and validated as with the previous SOX17 targeting. After a fully validated double-targeted clone (SOX17-tdTomato and BRA-mCerulean) was identified, it underwent sequential SOX2 targeting. Unlike in the cases of SOX17 and BRA, for SOX2 targeting, the direct fusion of mCitrine with SOX2 did not affect its localization or function, and therefore the SOX2-mCitrine reporter constitutes a faithful reporter of both the 'on' and 'off' expression rates. The SOX2 homology donor consists of: (i) a 1-kb left homology arm; (ii) an mCitrine-T2A-blasticidin cassette; and (iii) a 1-kb right homology arm. Colonies derived from single blasticidin-resistant cells were screened and validated as with the previous SOX17 and BRA targetings.

**RUES2-GLR time-lapse imaging.** Cultures of RUES2-GLR cells were dissociated to single cells from growth plates with StemPro Accutase (Life Technologies),

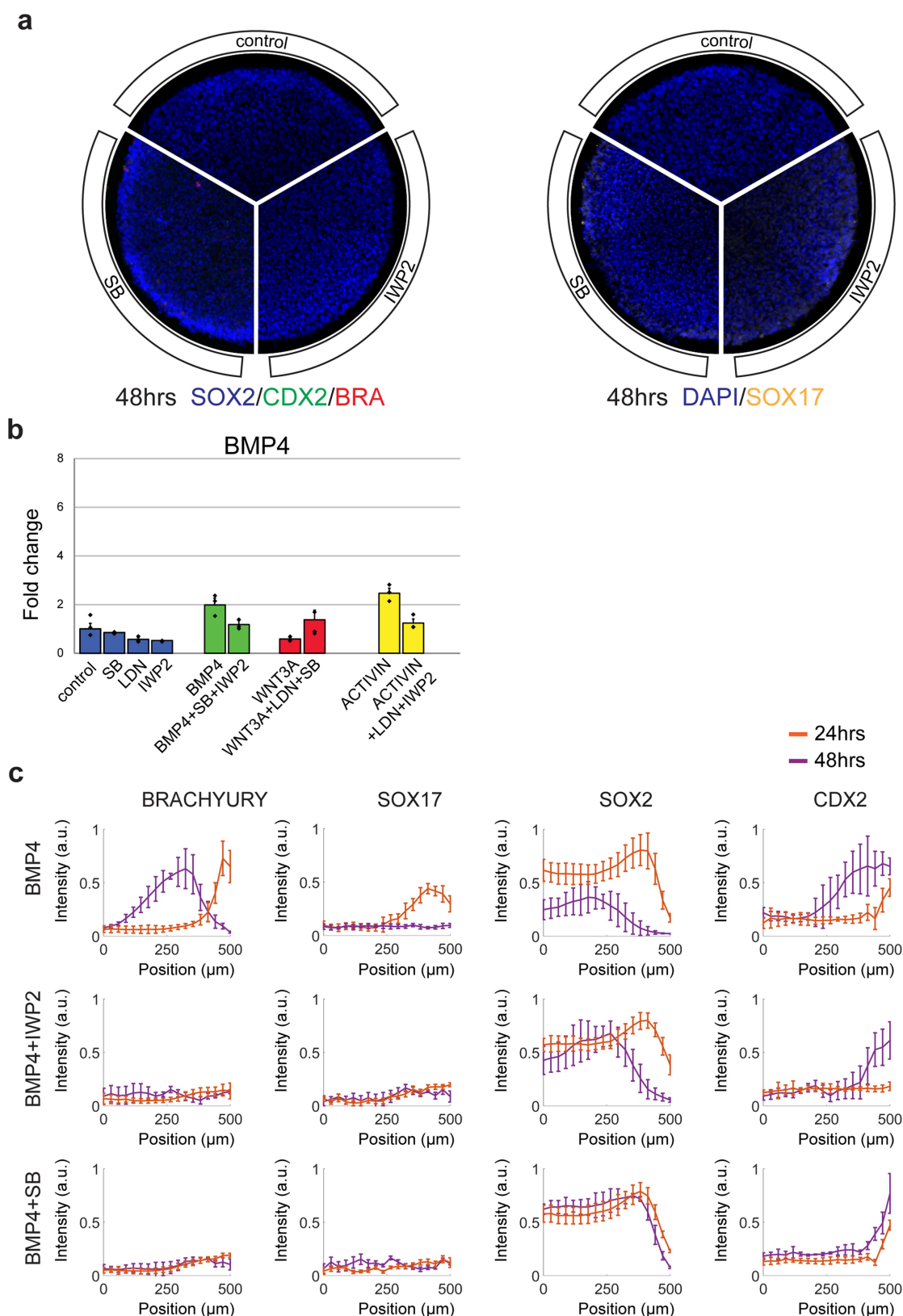
washed, and then re-suspended in MEF-CM with 10  $\mu$ M Y-27632 (Abcam). CYTOO micropatterned chips were placed in 35-mm tissue culture plastic dishes, and  $8 \times 10^5$  cells in 2 ml medium were added to each coverslip. After 1 h Y-27632 was removed and replaced with standard MEF-CM, supplemented with penicillin-streptomycin (Life Technologies), and incubated overnight. The following morning the micropatterned coverslip was carefully removed from the dish and placed in a coverslip holder (CYTOOchambers from CYTOO), to which 1 ml MEF-CM, penicillin-streptomycin, 50 ng/ml BMP4 was added to induce differentiation. Immediately after addition of medium, the holder was transferred to a spinning disk confocal microscope (CellVoyager CV1000, Yokogawa), in which fluorescent images were acquired every 30 min for 2 days. Multichannel time-lapse videos were generated from the raw images using ImageJ analysis software.

**Reporting summary.** Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

**Code availability statement.** Details or copies of the custom code are available upon request.

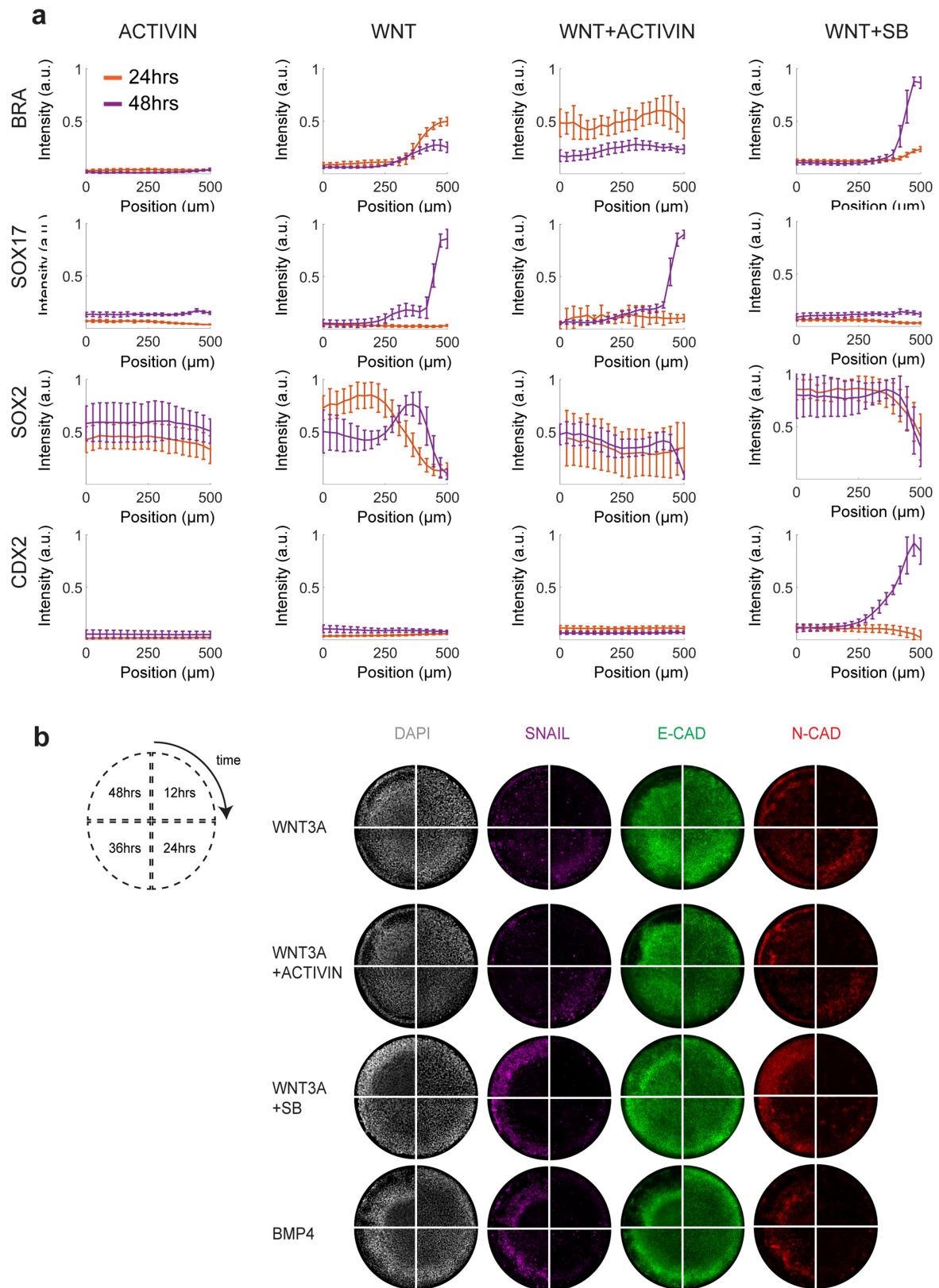
**Data availability.** All relevant data are available from the authors and the Source Data represented graphically in the figures are available in the online version of this paper. RNA-seq data are from a previously published dataset<sup>19</sup>, which is available from the GEO database under accession number GSE77057.

20. Vieceli, F. M. et al. The transcription factor chicken Scratch2 is expressed in a subset of early postmitotic neural progenitors. *Gene Expr. Patterns* **13**, 189–196 (2013).
21. Hamburger, V. & Hamilton, H. L. A series of normal stages in the development of the chick embryo. *J. Morphol.* **88**, 49–92 (1951).
22. Psychoyos, D. & Finnell, R. Assay for neural induction in the chick embryo. *J. Vis. Exp.* <https://doi.org/10.3791/1027> (2009).



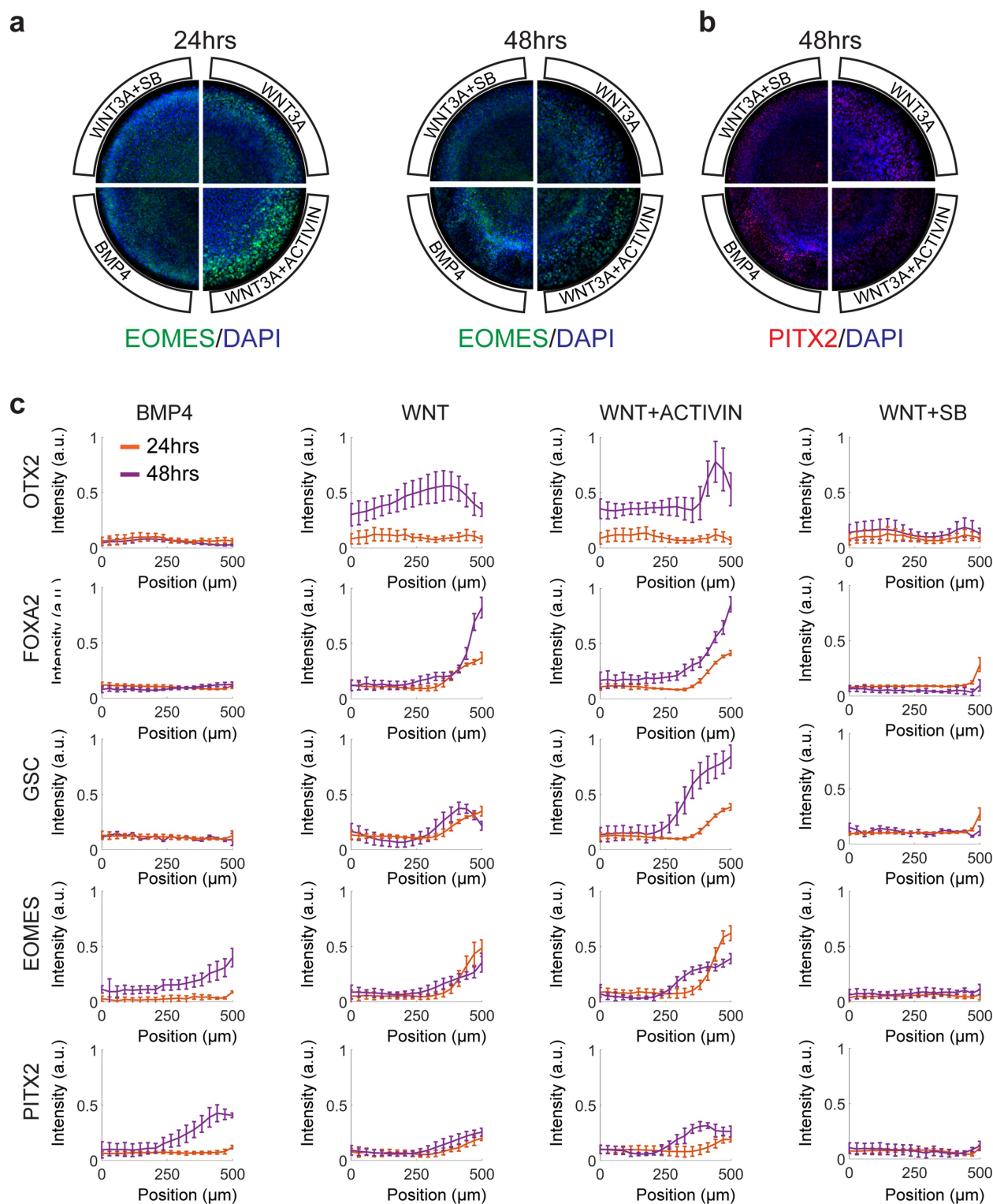
**Extended Data Fig. 1 | Controls for investigating hESC primitive streak initiation hierarchy.** **a**, Micropatterned hESC colonies were stimulated with IWP2, SB or blank medium. Colonies were fixed after 48 h and stained for germ layer molecular markers. This experiment was repeated at least three times independently with similar results. **b**, qPCR for BMP4 of unpatterned small colonies stimulated for 4 h as specified. Consistent with the model hierarchy, there was no marked induction of BMP4 by Activin,

WNT3A, or BMP4. Data are mean  $\pm$  s.d. of three biologically independent replicates. **c**, Quantification of expression data in Fig. 1d. Here, and in all other analyses unless stated otherwise, nuclei were segmented using DAPI and the intensity of immunofluorescence signal for each marker was normalized to the DAPI intensity. Single-cell expression data was binned radially and averaged. The final radial profile represents the mean  $\pm$  s.d. of  $n = 25$  colonies.



**Extended Data Fig. 2 | Primitive streak germ layer quantification and EMT timing.** **a**, Quantification of data presented in Fig. 2a. The radial profile represents the mean  $\pm$  s.d. of  $n = 25$  colonies. **b**, Micropatterned hESC colonies were stimulated with BMP4, WNT3A, WNT3A + SB, or WNT3A + Activin, fixed after 12, 24, 36 or 48 h and stained for the primitive streak molecular markers SNAIL, E-cadherin (E-CAD) and

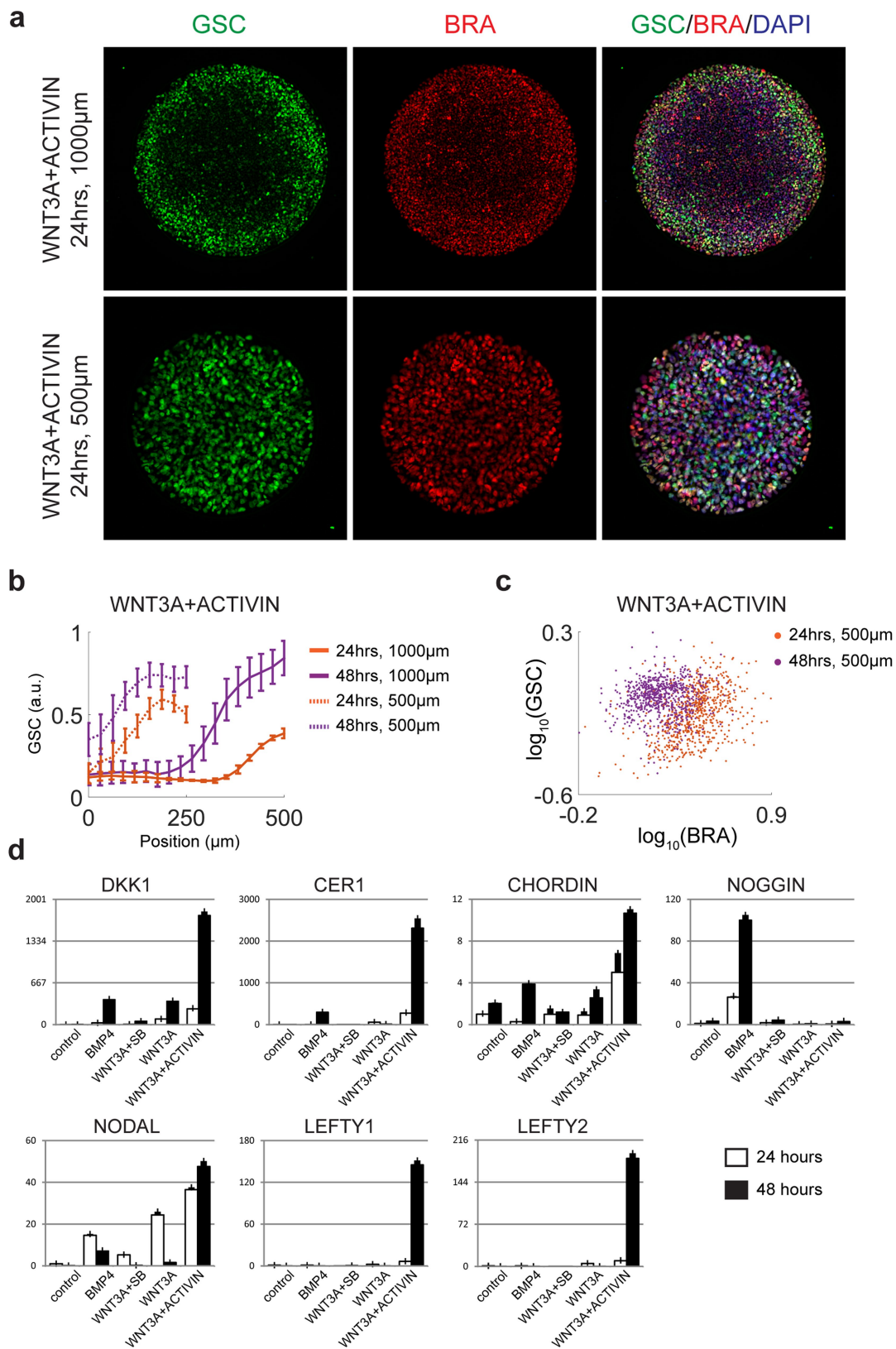
N-cadherin (N-CAD). Note that colonies stimulated with WNT3A and WNT3A + Activin turn on expression of EMT markers more rapidly than those stimulated with BMP4 or WNT3A + SB, and have mostly downregulated SNAIL after 48 h. This experiment was repeated at least three times independently with similar results.



### Extended Data Fig. 3 | Further micropattern fate characterization.

**a, b**, Micropatterned hESC colonies were stimulated with BMP4, WNT3A, WNT3A + SB, or WNT3A + Activin, fixed at 24 or 48 h after stimulation, and stained for EOMES (**a**) or PITX2 at 48 h (**b**). EOMES expression was highest in cells stimulated with WNT3A or WNT3A + Activin and was also dynamic, with the highest expression levels occurring 24 h

after stimulation, coinciding with the onset of primitive streak marker expression (Extended Data Fig. 2b). PITX2 is not highly expressed in any of the tested conditions. This experiment was repeated at least three times independently with similar results. **c**, Quantification of data in Fig. 2j, k and Extended Data Fig. 3a, b. The radial profile represents the mean  $\pm$  s.d. of  $n = 25$  colonies.

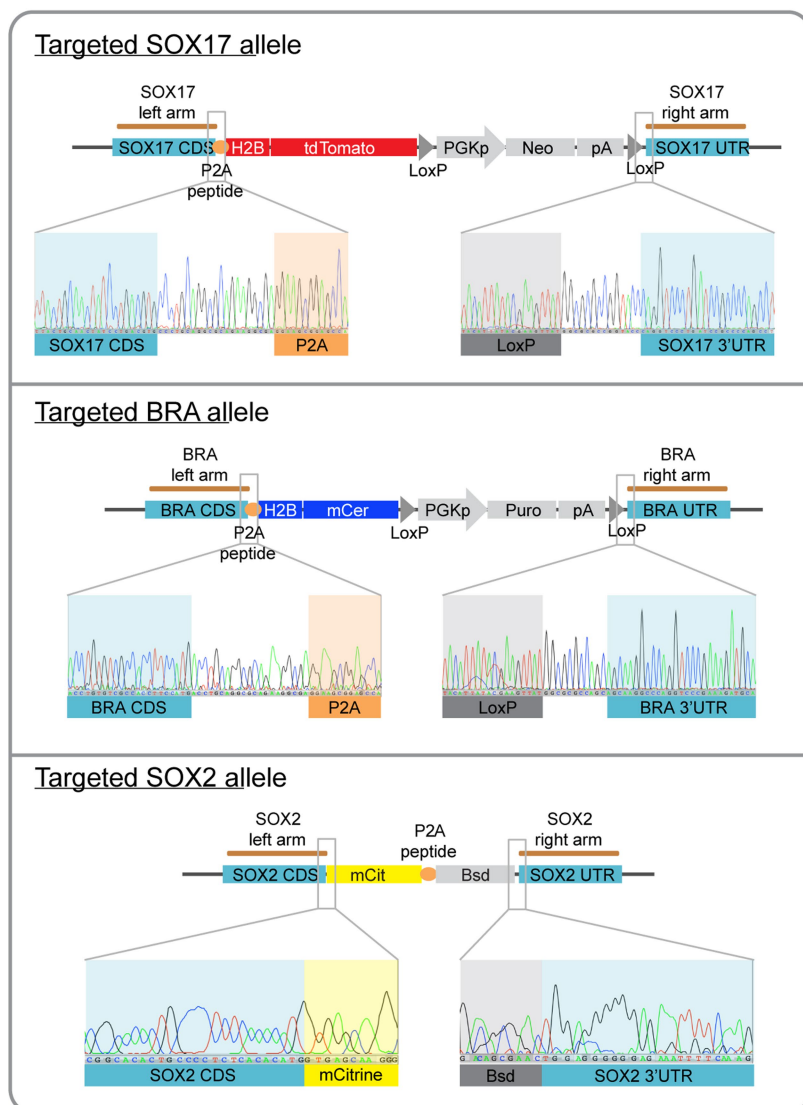
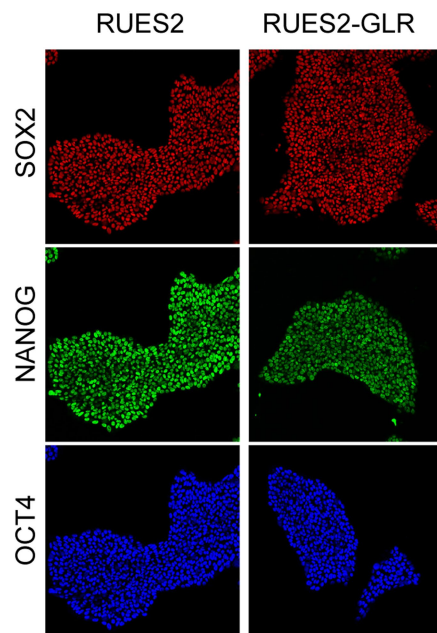
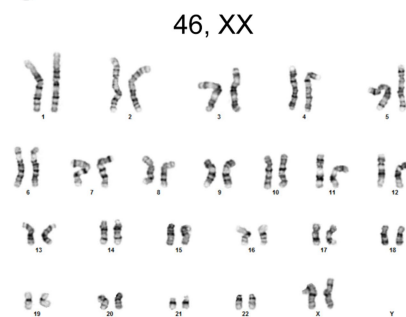


Extended Data Fig. 4 | See next page for caption.

**Extended Data Fig. 4 | Further characterization of the induced organizer.**

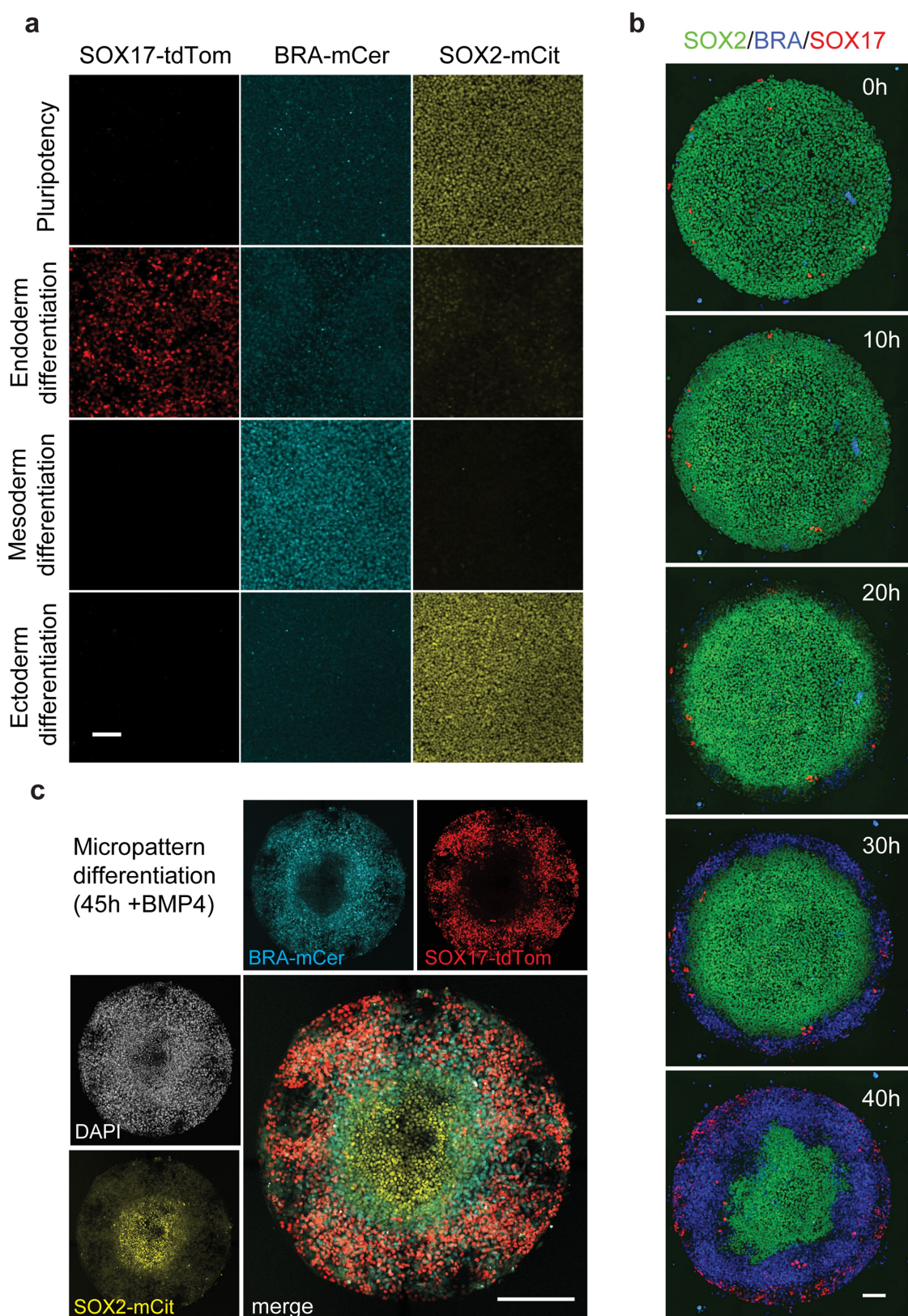
**a**, Micropatterned hESC colonies (1,000- $\mu\text{m}$  and 500- $\mu\text{m}$  diameter) stimulated with WNT3A + Activin, fixed 24 h after stimulation and stained for GSC and BRA. Note that, as previously observed<sup>1</sup> for BMP4 induction, shrinking the colony size results in removal of the central micropattern fate region, resulting here in a higher proportion of GSC-expressing cells. This experiment was repeated at least three times independently with similar results. **b**, Quantification of **a**. The radial profile represents the mean  $\pm$  s.d. of  $n = 25$  colonies. **c**, Scatter plot of single-cell expression of GSC versus BRA. Note that at 24 h most cells co-express BRA and GSC, but by 48 h GSC expression is increased and BRA expression is decreased. We therefore grafted micropatterns at 24 h as well as at 48 h post-stimulation, reasoning that earlier coexpression of BRA

and GSC would result in increased graft contribution to axial mesoderm structures. **d**, qPCR of additional organizer markers. RNA samples were collected from 500- $\mu\text{m}$ -diameter micropatterns stimulated with BMP4, WNT3A, WNT3A + SB, or WNT3A + Activin for 24 or 48 h. With the exception of *NOG*, the characteristic organizer-secreted inhibitors *DKK1*, *CER1*, *CHRD*, *LEFTY1* and *LEFTY2*, are all most highly expressed in the WNT3A + Activin condition. The high induction of *NOG* by BMP4 in hESCs has been noted before<sup>19</sup>, and may represent a species difference between human and mouse. *NODAL*, which in mouse is restricted to the organizer later in gastrulation, is also most highly expressed in WNT3A + Activin conditions. Data are mean  $\pm$  s.d. of three biologically independent replicates.

**a****b****c**

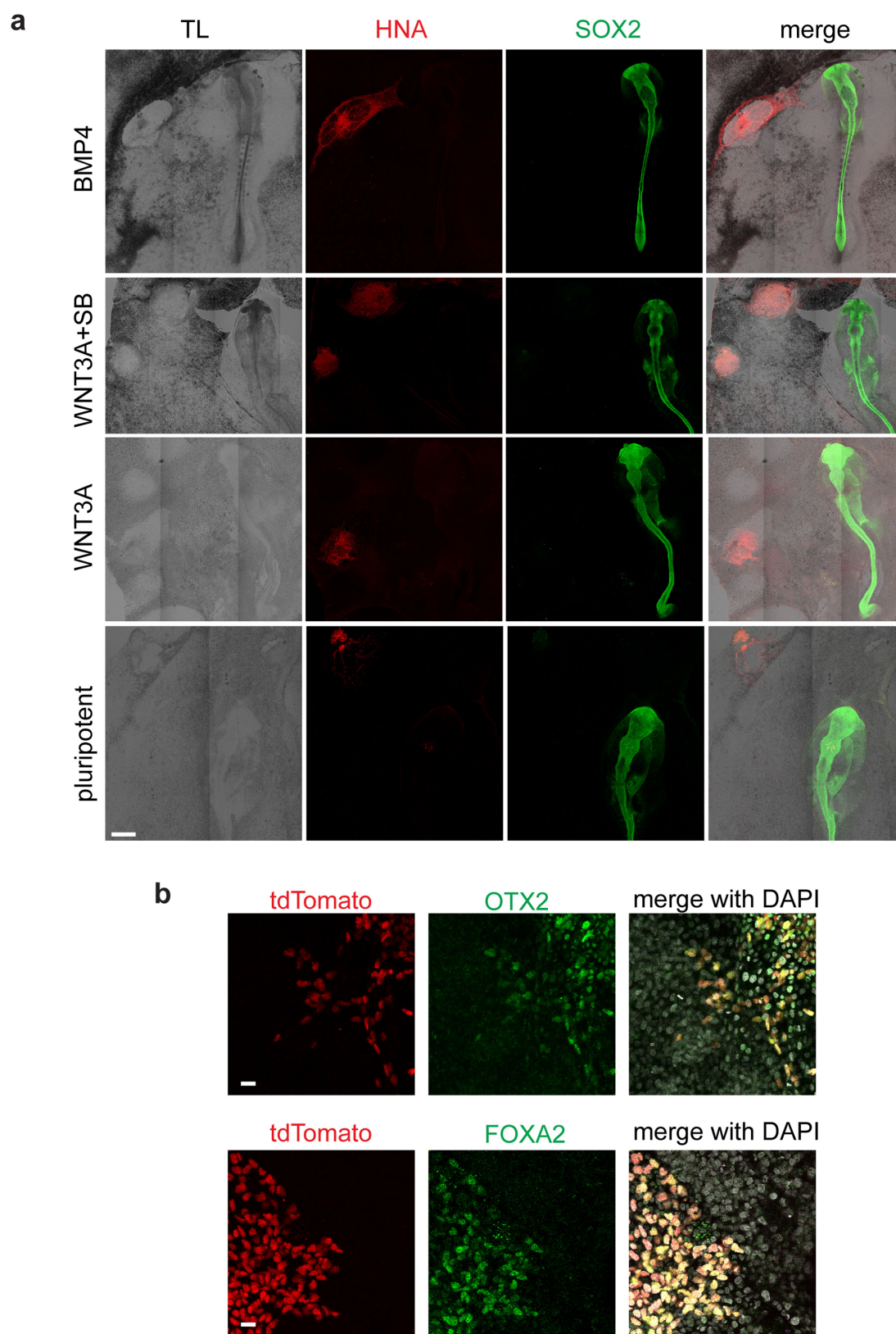
**Extended Data Fig. 5 | Generation and validation of the RUES2-GLR cell line. a,** Sequencing of the targeted alleles of *SOX17*, *BRA* and *SOX2* genes. No indels were detected. **b,** The RUES2-GLR cell line maintains pluripotency normally, as assessed by staining of typical pluripotency

markers (OCT4, NANOG and SOX2). Scale bar, 100  $\mu$ m. This experiment was repeated at least two times independently with similar results. **c,** The RUES2-GLR cell line was karyotypically normal.



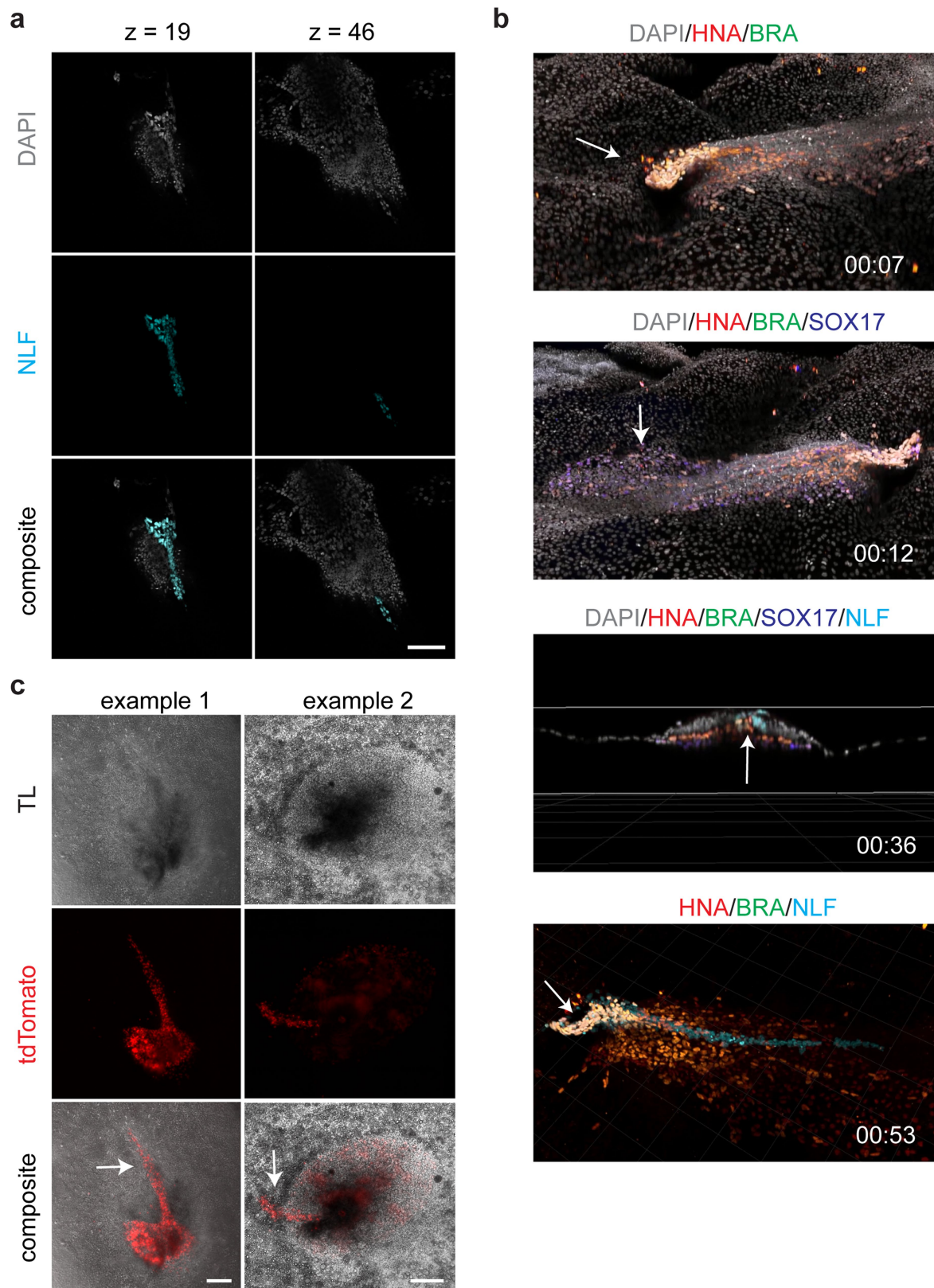
**Extended Data Fig. 6 | Functional validation of the RUES2-GLR cell line.** **a**, Specificity of germ layer reporters. When induced to differentiate to individual germ fates, only the specific reporter was turned on. SOX2-mCitrine was expressed during pluripotency and three days after neural (ectoderm) differentiation, BRA-mCerulean expression commenced after three days of mesodermal differentiation and SOX17-tdTomato reporter was active after three days of endodermal differentiation. Scale bar, 100  $\mu$ m. **b**, Snapshots of a time-lapse imaging of the RUES2-GLR

cells in micropatterns after treatment with 50 ng ml<sup>-1</sup> BMP4, showing how differentiation starts from the edges and extends inwards. Scale bar, 100  $\mu$ m. **c**, RUES2-GLR colonies reproducibly generate the typical self-organized concentric rings of germ layers when induced to differentiate with a step presentation of 50 ng ml<sup>-1</sup> BMP4 in micropatterns. Scale bar, 200  $\mu$ m. These experiments were repeated at least three times independently with similar results.



**Extended Data Fig. 7 | Control chick grafts.** **a**, Representative grafts for control conditions. With the exception of the BMP4 control condition, grafted hESC colonies were static, with the colonies either growing or dying in place. With BMP4 treatment, the colonies were frequently elongated, possibly owing to hESC migration. There was no induction of SOX2 in the host cells in any of the control conditions. Note that in the case of the WNT3A + SB graft shown, two colonies were grafted into

two different locations. Scale bar, 500  $\mu$ m. Experiments were repeated at least three times independently with similar results. **b**, Confocal sections showing co-expression of SOX17 (tdTomato) and FOXA2 or OTX2 in human cells that contribute to the secondary axes induced by a 24-h WNT3A + Activin-stimulated micropatterned hESC colony. Scale bar, 20  $\mu$ m. Experiments were repeated at least three times independently with similar results.



Extended Data Fig. 8 | See next page for caption.

**Extended Data Fig. 8 | Further characterization of the induced secondary axis. a,** Examples of classifying the notochord-like feature (NLF) based on morphology. At  $z = +19\mu\text{m}$ , the NLF appears as a tighter and brighter rod of cells running north to south that is also distinct and somewhat separated from the surrounding chick epiblast. At  $z = +46\mu\text{m}$ , paired elongated cells stick out ahead of the other cells in a continuation of the originally identified NLF. Other cells belonging to the NLF between  $z = +46\mu\text{m}$  and  $z = +19\mu\text{m}$  are obscured or out of focus in this section, but can be easily identified in individual sections at the other  $z$  positions. Scale bar,  $100\mu\text{m}$ . **b,** Snapshots from Supplementary Video 1. From top to bottom: yellow shows co-localization of BRA (green) and human

(red) cells; purple shows co-localization of Sox17:tdTomato (blue) with human (red) cells; cross-section shows that chick and human cells arrange themselves into germ layers properly, and that they flank the central notochord-like feature indicated by the arrow (cyan); a proportion of human mesoderm cells contribute to part of the notochord-like structure, while the cyan-coloured cells without HNA (red) shows that the remainder of the NLF is composed of host cells. **c,** Two examples of donor hESC grafts contributing to the induced notochord-like feature, imaged in a live chick embryo 27 h (left) and 23 h (right) after grafting. Scale bars,  $200\mu\text{m}$  (left),  $100\mu\text{m}$  (right). Similar notochord-like features were observed in at least ten independent biological replicates.

# Disruption of the beclin 1–BCL2 autophagy regulatory complex promotes longevity in mice

Álvaro F. Fernández<sup>1,2,9</sup>, Salwa Sebtí<sup>1,2,9</sup>, Yongjie Wei<sup>1,2,3</sup>, Zhongju Zou<sup>1,2,3</sup>, Mingjun Shi<sup>4</sup>, Kathryn L. McMillan<sup>4</sup>, Congcong He<sup>5</sup>, Tabitha Ting<sup>1,2</sup>, Yang Liu<sup>1,2,3</sup>, Wei-Chung Chiang<sup>1,2</sup>, Denise K. Marciano<sup>2</sup>, Gabriele G. Schiattarella<sup>2</sup>, Govind Bhagat<sup>6</sup>, Orson W. Moe<sup>2,4,7</sup>, Ming Chang Hu<sup>2,4\*</sup> & Beth Levine<sup>1,2,3,8\*</sup>

**Autophagy increases the lifespan of model organisms; however, its role in promoting mammalian longevity is less well-established<sup>1,2</sup>. Here we report lifespan and healthspan extension in a mouse model with increased basal autophagy. To determine the effects of constitutively increased autophagy on mammalian health, we generated targeted mutant mice with a Phe121Ala mutation in beclin 1 (*Becn1*<sup>F121A/F121A</sup>) that decreases its interaction with the negative regulator BCL2. We demonstrate that the interaction between beclin 1 and BCL2 is disrupted in several tissues in *Becn1*<sup>F121A/F121A</sup> knock-in mice in association with higher levels of basal autophagic flux. Compared to wild-type littermates, the lifespan of both male and female knock-in mice is significantly increased. The healthspan of the knock-in mice also improves, as phenotypes such as age-related renal and cardiac pathological changes and spontaneous tumorigenesis are diminished. Moreover, mice deficient in the anti-ageing protein *klotho*<sup>3</sup> have increased beclin 1 and BCL2 interaction and decreased autophagy. These phenotypes, along with premature lethality and infertility, are rescued by the beclin 1 (F121A) mutation. Together, our data demonstrate that disruption of the beclin 1–BCL2 complex is an effective mechanism to increase autophagy, prevent premature ageing, improve healthspan and promote longevity in mammals.**

Autophagy, an evolutionarily conserved lysosomal degradation pathway, has a key role in tissue homeostasis, health and disease<sup>4</sup>. In 2003, we showed that the *Caenorhabditis elegans* autophagy gene *bec-1* (orthologue of yeast *ATG6*, mammalian beclin 1) was required for lifespan extension in nematodes with a loss of function in the insulin signalling pathway<sup>5</sup>. Subsequently, numerous loss-of-function studies in *C. elegans* and *Drosophila* have confirmed an essential role for the autophagy machinery in longevity<sup>1,2</sup>, and tissue-specific deletion of core autophagy genes have shown that autophagy delays age-related changes in mouse tissues, including kidney and heart<sup>6,7</sup>. Moreover, physiological inducers (such as caloric restriction) as well as pharmacological inducers (such as spermidine) of autophagy increase lifespan in mice<sup>1,8,9</sup>. Despite these clues that autophagy may be a longevity pathway in mammals, definitive evidence that increased basal autophagy extends mammalian healthspan and lifespan is lacking.

An earlier study<sup>10</sup> demonstrated an increase in lifespan of mice that transgenically overexpress *ATG5*. However, it is unclear how overexpression of *ATG5*, a protein necessary for autophagy but not directly involved in the regulation of autophagy levels, results in increased autophagy. Moreover, *ATG5* has other key functions, such as the regulation of inflammation<sup>11</sup>, and these roles are not shared by other genes in the autophagy pathway. Therefore, it is imperative to use a more direct and specific genetic approach to assess the effects of enhanced basal autophagy on mammalian lifespan and healthspan. To do so,

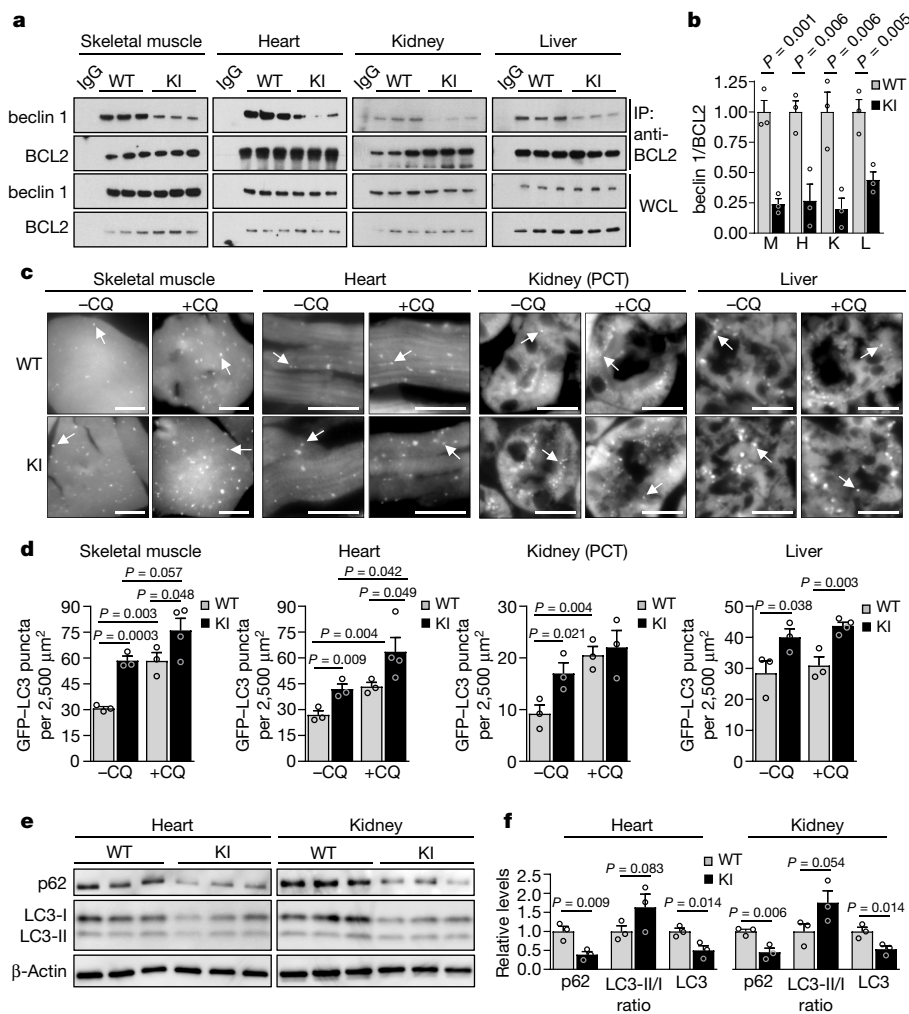
we focused on the mammalian autophagy protein, beclin 1 (encoded by *Becn1*)<sup>12</sup>, which is part of an autophagy-specific class III phosphatidylinositol-3-OH kinase (PI3K) complex<sup>13</sup> that has a key role in the regulation of the initiation of autophagosome formation<sup>14</sup>.

We recently reported the construction of mice with a Phe-to-Ala knock-in substitution mutation in the BH3 domain of beclin 1 (F121A; corresponding to F123A in human beclin 1)<sup>15</sup> that decreases the binding of two negative regulators of autophagy (BCL2 and BCL-XL) to beclin 1 in vitro<sup>16,17</sup>. Using these mice, we performed co-immunoprecipitation of endogenous beclin 1 with BCL2 in muscle, heart, kidney and liver of two-month-old wild-type and homozygous knock-in mice. We observed a marked reduction in beclin 1 co-immunoprecipitation with BCL2 in the tissues of the knock-in mice (Fig. 1a, b). In parallel, we analysed autophagic flux by crossing wild-type or knock-in mice with animals that transgenically express green fluorescent protein (GFP)-tagged LC3<sup>18</sup>, a fluorescent marker of autophagosomes. In skeletal muscle, heart, renal glomeruli, proximal convoluted tubules and liver, knock-in mice had significantly increased numbers of GFP–LC3 puncta compared to wild-type control littermates (Fig. 1c, d; Extended Data Fig. 1). In all tissues except for the liver, there was a further increase in GFP–LC3 puncta after treatment with chloroquine, an inhibitor of lysosomal acidification and autophagosome–lysosomal fusion, indicating that the increased numbers of GFP–LC3 puncta in knock-in mice represents a true increase in basal autophagic flux, rather than a block in autophagosomal maturation. We further confirmed that knock-in mice had increased autophagic flux by western blot analyses. Both hearts and kidneys had increased conversion of LC3-I to LC3-II (the lipidated, autophagosome-associated form of LC3), decreased levels of total LC3 and decreased levels of the autophagy substrate p62<sup>19</sup> (Fig. 1e, f). Similar findings were also observed in the hearts and kidneys of six- to eight-month-old mice (Extended Data Fig. 2), indicating that the effects of the knock-in mutation are sustained over time in adulthood.

We further evaluated the effect of the knock-in mutation on autophagy using mouse embryonic fibroblasts (MEFs) derived from knock-in or wild-type littermate controls. In knock-in MEFs, there was decreased co-immunoprecipitation of beclin 1 with BCL2, increased numbers of GFP–LC3 puncta, decreased levels of p62 and total LC3, and increased numbers of autophagic structures (autophagosomes and autolysosomes) observed by quantitative electron microscopy (Extended Data Fig. 3a–d). Treatment with the lysosomal inhibitor bafilomycin A1 indicated a bona fide increase in autophagic flux in knock-in MEFs. To evaluate possible effects of the beclin 1 knock-in mutation other than autophagy, we examined endocytosis, a process thought to involve the beclin 1–VPS34 complex<sup>20</sup>. We did not observe any significant differences in endocytosis between knock-in and

<sup>1</sup>Center for Autophagy Research, University of Texas Southwestern Medical Center, Dallas, TX, USA. <sup>2</sup>Department of Internal Medicine, University of Texas Southwestern Medical Center, Dallas, TX, USA. <sup>3</sup>Howard Hughes Medical Institute, University of Texas Southwestern Medical Center, Dallas, TX, USA. <sup>4</sup>Charles and Jane Pak Center for Mineral Metabolism and Clinical Research, University of Texas Southwestern Medical Center, Dallas, TX, USA. <sup>5</sup>Department of Cell and Molecular Biology, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA. <sup>6</sup>Department of Pathology and Cell Biology, Columbia University Medical Center and New York Presbyterian Hospital, New York, NY, USA. <sup>7</sup>Department of Physiology, University of Texas Southwestern Medical Center, Dallas, TX, USA. <sup>8</sup>Department of Microbiology, University of Texas Southwestern Medical Center, Dallas, TX, USA. <sup>9</sup>These authors contributed equally: Álvaro F. Fernández, Salwa Sebtí.

\*e-mail: ming-chang.hu@utsouthwestern.edu; beth.levine@utsouthwestern.edu

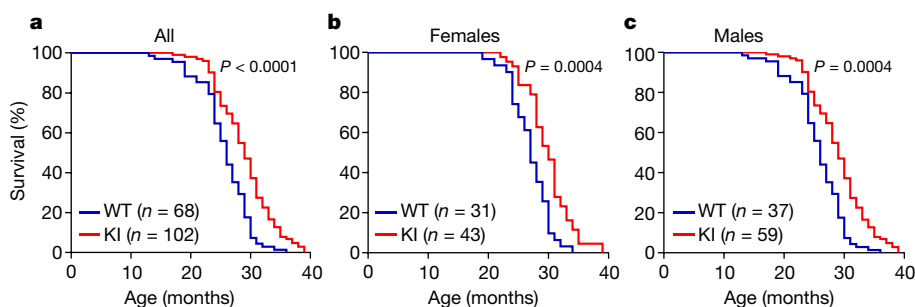


**Fig. 1 | Effects of beclin 1(F121A) mutation on the beclin 1-BCL2 interaction and basal autophagy.** **a**, Co-immunoprecipitation of beclin 1 and BCL2 in indicated tissues from two-month-old *Becn1*<sup>+/+</sup> (wild-type, WT) and *Becn1*<sup>F121A/F121A</sup> (knock-in, KI) animals. **b**, Quantification of beclin 1 co-immunoprecipitated with BCL2 in **a**. H, heart; K, kidney; L, liver; M, skeletal muscle. **c**, Representative images of GFP-LC3 puncta (autophagosomes) in indicated tissues from wild-type and knock-in mice that had been crossed with mice that transgenically express GFP-LC3, with or without 50 mg kg<sup>-1</sup> chloroquine (CQ) for 6 h. Scale bars, 10  $\mu\text{m}$ . (See Extended Data Fig. 1 for enlarged images.) White arrows denote

representative GFP-LC3 puncta. **d**, Quantification of GFP-LC3 puncta with or without chloroquine in indicated tissues. **e**, Western blot analysis of autophagy markers in the hearts and kidneys of two-month-old wild-type and knock-in mice. **f**, Quantification of p62 and total LC3 levels (normalized to  $\beta$ -actin) and LC3-II/LC3-I ratio in **e**. Data are mean  $\pm$  s.e.m. for three mice per genotype. In **a** and **e**, each lane represents a different mouse. *P* values were determined by a one-sided unpaired *t*-test. PCT, renal proximal convoluted tubules; WCL, whole cell lysate. For uncropped gels, see Supplementary Fig. 1.

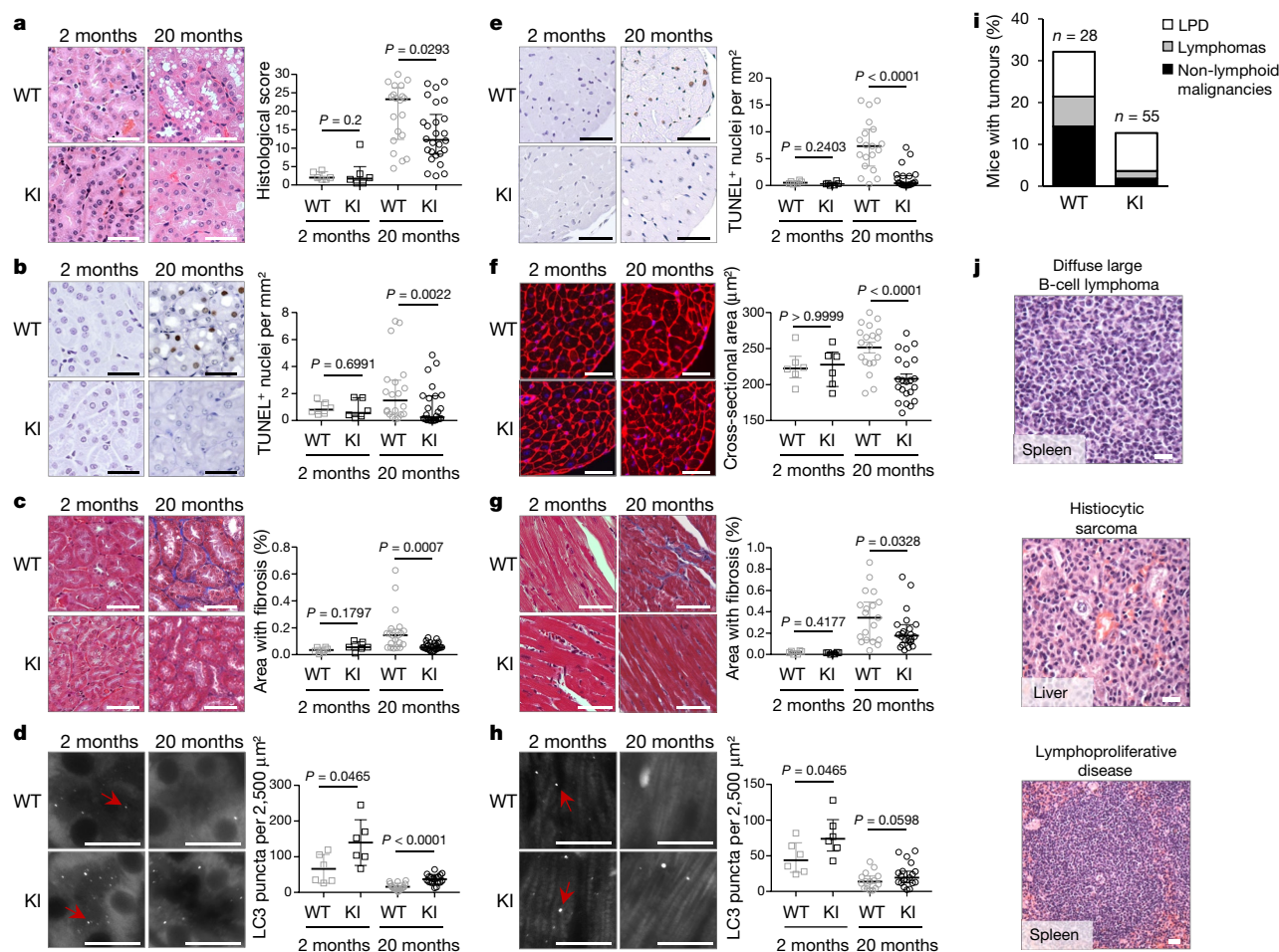
wild-type MEFs, as measured by the kinetics of endocytic uptake of fluorescent transferrin (Extended Data Fig. 3e). This is expected, as BCL2 binding to beclin 1 has not been shown to regulate beclin 1-dependent functions other than autophagy.

Taken together, these findings demonstrate that the *Becn1*<sup>F121A/F121A</sup> knock-in mouse model is useful for studying the effects of constitutively increased basal autophagy on mammalian lifespan and healthspan. Therefore, we evaluated the lifespan of a large cohort



**Fig. 2 | Beclin 1(F121A) knock-in mutation extends lifespan in mice.** **a–c**, Kaplan–Meier survival curves for *Becn1*<sup>+/+</sup> wild-type and *Becn1*<sup>F121A/F121A</sup> knock-in mice, showing the lifespan of all mice in the

cohort (**a**), females alone (**b**) or males alone (**c**). *n* denotes the number of mice per group. *P* values were determined by log-rank (Mantel–Cox) test.



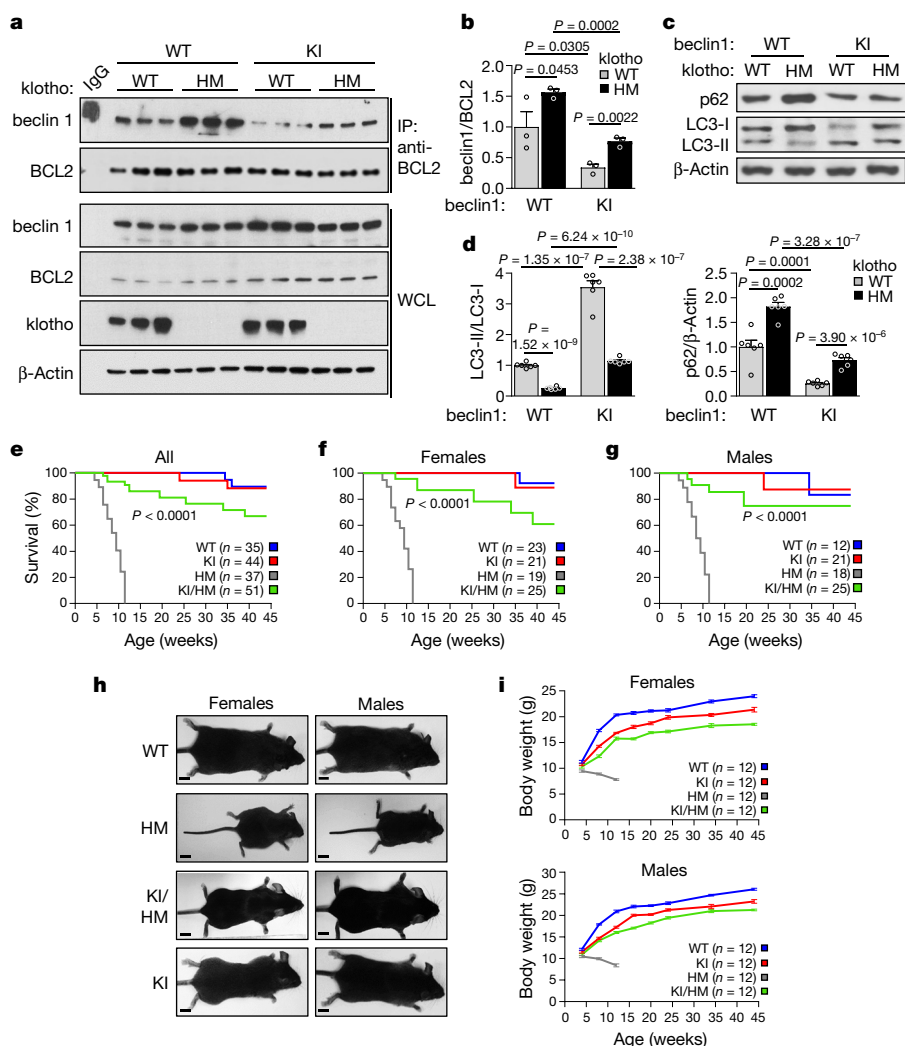
**Fig. 3 | Beclin 1(F121A) knock-in mutation improves healthspan in mice.** **a–d**, Representative images and quantification of pathological score (**a**), TUNEL-positive nuclei (**b**), interstitial fibrosis (**c**), and endogenous LC3 puncta (autophagosomes) (see enlarged images in Extended Data Fig. 4c) (**d**) in the cortical region of the kidney. **e–h**, Representative images and quantification of TUNEL-positive nuclei (**e**), cardiomyocyte cross-sectional fibre size (**f**), cardiac interstitial fibrosis (**g**), and endogenous LC3 puncta (see enlarged images in Extended Data Fig. 4d) (**h**) in the heart. Two-month-old wild-type and knock-in mouse kidneys and hearts ( $n = 6$  per genotype) were analysed. For kidney analyses, 20-month-old wild-type ( $n = 20$  and  $n = 16$ ) and knock-in ( $n = 26$  and  $n = 19$ ) mice were used for histopathological and autophagy analyses, respectively. For heart analyses,

20-month-old wild-type ( $n = 19$  and  $n = 15$ ) and knock-in ( $n = 22$  and  $n = 19$ ) mice were used for histopathological and autophagy analyses, respectively. Scatter plot bars represent median  $\pm$  interquartile ranges.  $P$  values were determined by Mann–Whitney test. Autophagy analyses were one-sided and all other analyses were two-sided. **i, j**, Percentage of wild-type and knock-in mice (aged 20 months) with spontaneous tumours, including lymphoproliferative disease (LPD), lymphomas, and non-lymphoid malignancies (**i**) and representative images of the most frequently observed neoplastic lesions from wild-type mice (**j**). See text for statistical analyses of data. Scale bars, 50  $\mu\text{m}$  (**a–c**, **e–g**, **j**) and 10  $\mu\text{m}$  (**d**, **h**). In **d** and **h**, red arrows denote representative endogenous LC3 puncta.

of beclin 1 knock-in and wild-type littermates (Fig. 2) on an inbred C57BL/6J background. The combined data for males and females showed a significant lifespan extension (Fig. 2a) of knock-in mice compared to wild-type littermate controls (median survival: 26 months, wild type; 29 months, knock-in). This extension in longevity was observed both in females (median survival: 27 months, wild type; 30 months, knock-in) and males (median survival: 25 months, wild type; 28 months, knock-in) (Fig. 2b, c; Extended Data Table 1), showing a sex-independent effect of the beclin 1(F121A) mutation on lifespan. Knock-in mice also had an increase in maximal lifespan (Extended Data Table 1). Thus, this gain-of-function mutation in a core autophagy gene, *Becn1*, extends mammalian lifespan.

Next, we evaluated age-related phenotypes in two vital organs, kidney and heart, where we showed that the beclin 1(F121A) mutation disrupted beclin 1–BCL2 binding and increased basal autophagy. During ageing, the kidney develops pathological changes, including fibrosis and nuclear damage, and proximal convoluted tubule-specific deletion of the autophagy gene, *Atg5*, exacerbates these phenotypes<sup>6,21</sup>. In 20-month-old knock-in mice and wild-type control littermates, the proximal convoluted tubules had increased vacuolar changes in

the wild-type mice as compared to the knock-in mice (Fig. 3a). In this region, there was increased nuclear DNA damage, as measured by increased TUNEL staining of non-pyknotic/apoptotic-appearing nuclei (Fig. 3b) that did not stain positive for the apoptotic marker, active caspase 3 (Extended Data Fig. 4a). Consistent with this evidence of decreased cellular damage in the proximal convoluted tubules of knock-in mice, there was also decreased fibrosis (Fig. 3c). Similarly, we observed increased non-pyknotic/apoptotic-appearing, active caspase 3-negative, TUNEL-positive nuclei in the hearts of 20-month-old wild-type as compared to knock-in mice (Fig. 3e; Extended Data Fig. 4b). This phenotype was most pronounced in the subendocardial region of the left ventricle (the part of the myocardium subjected to greatest haemodynamic stress), and was accompanied by an increase in the cardiomyocyte cross-sectional area in the wild-type mice (Fig. 3f). Moreover, there was a significant increase in fibrosis, a hallmark of cardiac ageing, in wild-type versus knock-in mice (Fig. 3g). In young (two-month-old) mice, no genotype-specific differences were observed in any of the renal and cardiac parameters assessed, indicating that the decreased pathology in older knock-in mice genuinely reflects a reduction in age-related changes (rather than developmental differences) in



**Fig. 4 | Expression of beclin 1 (F121A) prevents lethality of klotho-deficient mice.**

**a**, Co-immunoprecipitation of beclin 1 and BCL2 in the kidneys from mice of indicated genotype ( $n = 3$  animals per genotype). **b**, Quantification of beclin 1 co-immunoprecipitated with BCL2 in **a**. **c**, Western blot analysis of autophagy markers in kidney from mice of indicated genotype. **d**, Quantification of western blot analysis of p62 levels and LC3-II/LC3-I ratios (combined results are for six mice per genotype; each independent experiment had one mouse per genotype). Data are mean  $\pm$  s.e.m. **e–g**, Kaplan–Meier survival curves for mice of indicated genotype, showing the lifespan of all the mice in the cohort (**e**), females alone (**f**) or males alone (**g**).  $n$  denotes the number of mice per group. **h**, Representative images of ten-week-old mice of indicated genotype for body size comparison. Scale bars, 1 cm. **i**, Serial body weights in female and male mice of indicated genotype. Data are mean  $\pm$  s.e.m. for 12 mice per genotype.  $P$  values were determined by one-sided unpaired  $t$ -test (**b**, **d**) or log-rank Mantel–Cox test (**e–g**). HM,  $Kl^{HM/HM}$  (hypomorphic mutation) mice; KI/HM,  $Becn1^{F121A/F121A} Kl^{HM/HM}$  mice. For uncropped gels, see Supplementary Fig. 1.

these organs. Thus, the beclin 1(F121A) knock-in mutation decreases age-related renal and cardiac changes, including fibrosis.

In this cohort of 20-month-old mice, autophagosome numbers, as measured by endogenous staining of LC3, were markedly decreased in the kidneys and hearts of both wild-type and knock-in mice (Fig. 3d, h; Extended Data Fig. 4c, d), consistent with a predicted age-related decline in autophagic capacity<sup>2,22</sup>. However, the knock-in mouse kidneys and hearts still had more autophagosomes than their wild-type counterparts. Taken together, these data demonstrate that the beclin 1(F121A) knock-in mutation delays, but does not prevent, age-related decline in mouse autophagic function in parallel with observed decreases in age-related organ pathology. This finding is predicted, as downstream autophagy gene expression decreases with ageing<sup>2,22</sup>; thus, decreased negative regulation of beclin 1 per se is not sufficient to reverse age-related decline in autophagosome formation.

In mice and humans, ageing results in increased susceptibility to a variety of cancers. Specifically, mice with a C57BL/6J genetic background display age-related increases in lymphoproliferative disease (LPD) and lymphomas, histiocytic sarcomas, lung cancers and liver cancers<sup>23</sup>. In a cohort of 20-month-old knock-in mice and wild-type littermates, there was a significant decrease in age-related spontaneous tumorigenesis in the knock-in mice (Fig. 3i, j), either when comparing all malignancies ( $P = 0.034$ , chi-square) or non-lymphoid malignancies alone ( $P = 0.024$ , chi-square). Thus, beclin 1(F121A) knock-in mice with increased basal autophagy have a decreased incidence of age-related spontaneous cancer. The similar prevalence of LPD in both genotypes, coupled with increased prevalence of lymphomas in

wild-type mice, is consistent with delayed onset of LPD and/or progression of LPDs to frank lymphoma in knock-in animals.

The central role of beclin 1–BCL2 disruption in promoting increased basal autophagy, lifespan extension and improved healthspan raised the question of whether known anti-ageing factors might exert their longevity-promoting activity by disrupting beclin 1–BCL2 binding and autophagy. Klotho is a single-pass membrane-bound protein that can be cleaved and released into the circulation; it is highly expressed in the kidney, and it regulates ageing<sup>24</sup>. Klotho expression decreases with ageing in mice and humans<sup>24</sup>. Klotho deficiency in mice, either via hypomorphic expression or gene knockout, results in a syndrome resembling (but not completely recapitulating) ageing<sup>3,25</sup>, including premature lethality (death within a few months of birth), infertility and multisystem defects. Klotho overexpression or administration of soluble klotho protein extends lifespan and rescues most phenotypes in klotho hypomorphic mice<sup>26,27</sup>. Moreover, soluble klotho promotes autophagic flux in the kidney and, in the setting of renal ischaemic injury, mitigates renal fibrosis and retards progression to end-stage renal disease<sup>28</sup>.

Therefore, we crossed the  $Becn1^{F121A/F121A}$  mice with a well-characterized premature ageing model in which animals harbour a hypomorphic mutation in the klotho gene<sup>3</sup> ( $kl/kl$  mice; termed here  $Kl^{HM/HM}$ ). Klotho deficiency resulted in a marked increase in beclin 1 co-immunoprecipitation with BCL2 in the kidney (Fig. 4a, b), in support of the concept that disruption of beclin 1 and BCL2 binding may have a mechanistic role in klotho-induced autophagy. Consistent with this, recombinant soluble klotho decreased beclin 1 and BCL2 binding in cultured HeLa cells (Extended Data Fig. 5). The beclin 1(F121A)

mutation decreased beclin 1 and BCL2 binding in  $Kl^{HM/HM}$  kidney to levels similar to those observed in wild-type mice without discernible effects on klotho protein expression (Fig. 4a, b). In parallel, the beclin 1 (F121A) mutation restored levels of autophagy in  $Kl^{HM/HM}$  mice to those observed in wild-type mice, as measured by LC3-II conversion and p62 degradation (Fig. 4c, d). Notably, this reversal of the enhanced beclin 1–BCL2 binding and decreased autophagy in the beclin 1 knock-in/ $Kl^{HM/HM}$  double-mutant mice nearly completely rescued the premature lethality phenotype of both female and male mice (Fig. 4e–g); 100% of the  $Kl^{HM/HM}$  mice were dead by approximately 12 weeks, whereas most of the double-mutant beclin 1 knock-in and  $kl/kl$  mice ( $Becn1^{F121A/F121A} Kl^{HM/HM}$ ) survived a 45-week observation period. Furthermore, both female and male infertility was rescued in the double-mutant ( $Becn1^{F121A/F121A} Kl^{HM/HM}$ ) mice. Finally, the severe growth retardation of  $Kl^{HM/HM}$  mice was almost completely reversed by the beclin 1 (F121A) mutation (Fig. 4h, i). Taken together, these data indicate that a gain-of-function mutation in beclin 1, which disrupts beclin 1–BCL2 binding and increases autophagy, reverses the premature ageing phenotypes resulting from klotho deficiency.

Our findings demonstrate that a knock-in gain-of-function point mutation in a key autophagy regulatory gene, *Becn1*, in mice results in increased basal levels of autophagy, extends lifespan in both males and females, and improves healthspan, including a decrease in age-related renal and cardiac changes and age-related spontaneous malignancies. Importantly, the decreased beclin 1–BCL2 binding in vivo is not only associated with longevity and improved ageing-related phenotypes, but it also has a marked effect on rescuing the early lethality and infertility of mice deficient in a well-established anti-ageing protein, klotho. Taken together, our results suggest that activation of the beclin 1 class III PI3K autophagy-initiating complex may be an effective and safe way to bypass upstream ageing signals to promote mammalian healthspan and lifespan. Specifically, we propose that disruption of beclin 1–BCL2 binding may be one such strategy, as our findings provide genetic proof that chronic in vivo disruption of this complex exerts substantial beneficial effects on mammalian lifespan and healthspan. More speculatively, it is possible that disruption of the beclin 1–BCL2 complex, and ensuing autophagy induction, underlies the anti-ageing mechanism of klotho and perhaps other longevity signalling pathways.

## Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0162-7>.

Received: 10 September 2017; Accepted: 16 April 2018;

Published online: 30 May 2018

- Madeo, F., Zimmermann, A., Maiuri, M. C. & Kroemer, G. Essential role for autophagy in life span extension. *J. Clin. Invest.* **125**, 85–93 (2015).
- Rubinsztein, D. C., Mariño, G. & Kroemer, G. Autophagy and aging. *Cell* **146**, 682–695 (2011).
- Kuro-o, M. et al. Mutation of the mouse *klotho* gene leads to a syndrome resembling ageing. *Nature* **390**, 45–51 (1997).
- Levine, B. & Kroemer, G. Autophagy in the pathogenesis of disease. *Cell* **132**, 27–42 (2008).
- Meléndez, A. et al. Autophagy genes are essential for dauer development and life-span extension in *C. elegans*. *Science* **301**, 1387–1391 (2003).
- Yamamoto, T. et al. Time-dependent dysregulation of autophagy: Implications in aging and mitochondrial homeostasis in the kidney proximal tubule. *Autophagy* **12**, 801–813 (2016).
- Eisenberg, T. et al. Cardioprotection and lifespan extension by the natural polyamine spermidine. *Nat. Med.* **22**, 1428–1438 (2016).
- Eisenberg, T. et al. Induction of autophagy by spermidine promotes longevity. *Nat. Cell Biol.* **11**, 1305–1314 (2009).
- Mercken, E. M. et al. SIRT1 but not its increased expression is essential for lifespan extension in caloric-restricted mice. *Ageing Cell* **13**, 193–196 (2014).

- Pyo, J. O. et al. Overexpression of Atg5 in mice activates autophagy and extends lifespan. *Nat. Commun.* **4**, 2300 (2013).
- Kimmy, J. M. et al. Unique role for ATG5 in neutrophil-mediated immunopathology during *M. tuberculosis* infection. *Nature* **528**, 565–569 (2015).
- Liang, X. H. et al. Induction of autophagy and inhibition of tumorigenesis by beclin 1. *Nature* **402**, 672–676 (1999).
- Kihara, A., Kabeya, Y., Ohsumi, Y. & Yoshimori, T. Beclin-phosphatidylinositol 3-kinase complex functions at the trans-Golgi network. *EMBO Rep.* **2**, 330–335 (2001).
- Levine, B., Liu, R., Dong, X. & Zhong, Q. Beclin orthologs: integrative hubs of cell signaling, membrane trafficking, and physiology. *Trends Cell Biol.* **25**, 533–544 (2015).
- Rocchi, A. et al. A *Becn1* mutation mediates hyperactive autophagic sequestration of amyloid oligomers and improved cognition in Alzheimer's disease. *PLoS Genet.* **13**, e1006962 (2017).
- Pattingre, S. et al. Bcl-2 antiapoptotic proteins inhibit Beclin 1-dependent autophagy. *Cell* **122**, 927–939 (2005).
- Sinha, S., Colbert, C. L., Becker, N., Wei, Y. & Levine, B. Molecular basis of the regulation of Beclin 1-dependent autophagy by the  $\gamma$ -herpesvirus 68 Bcl-2 homolog M11. *Autophagy* **4**, 989–997 (2008).
- Mizushima, N., Yamamoto, A., Matsui, M., Yoshimori, T. & Ohsumi, Y. *In vivo* analysis of autophagy in response to nutrient starvation using transgenic mice expressing a fluorescent autophagosome marker. *Mol. Biol. Cell* **15**, 1101–1111 (2004).
- Mizushima, N., Yoshimori, T. & Levine, B. Methods in mammalian autophagy research. *Cell* **140**, 313–326 (2010).
- McKnight, N. C. et al. Beclin 1 is required for neuron viability and regulates endosome pathways via the UVRAG-VPS34 complex. *PLoS Genet.* **10**, e1004626 (2014).
- Lim, J. H. et al. Age-associated molecular changes in the kidney in aged mice. *Oxid. Med. Cell. Longev.* **2012**, 171383 (2012).
- Lapierre, L. R., Kumsta, C., Sandri, M., Ballabio, A. & Hansen, M. Transcriptional and epigenetic regulation of autophagy in aging. *Autophagy* **11**, 867–880 (2015).
- Brayton, C. in *The Mouse in Biomedical Research* Vol. 2 (eds Fox, J. et al.), 623–717 (Elsevier, Amsterdam, 2007).
- Kuro-o, M. Klotho and aging. *Biochim. Biophys. Acta* **1790**, 1049–1058 (2009).
- Tsujioka, H., Kurotaki, Y., Fujimori, T., Fukuda, K. & Nabeshima, Y. *Klotho*, a gene related to a syndrome resembling human premature aging, functions in a negative regulatory circuit of vitamin D endocrine system. *Mol. Endocrinol.* **17**, 2393–2403 (2003).
- Kurosuo, H. et al. Suppression of aging in mice by the hormone Klotho. *Science* **309**, 1829–1833 (2005).
- Chen, T. H. et al. The secreted Klotho protein restores phosphate retention and suppresses accelerated aging in Klotho mutant mice. *Eur. J. Pharmacol.* **698**, 67–73 (2013).
- Shi, M. et al.  $\alpha$ Klotho mitigates progression of AKI to CKD through activation of autophagy. *J. Am. Soc. Nephrol.* **27**, 2331–2345 (2016).

**Acknowledgements** We thank S. Sciarretta for discussions; N. Mizushima for reagents; L. Nguyen for technical assistance and H. Smith for assistance with manuscript preparation. This work was supported by NIH grants R01-CA109618 (B.L.), U19-AI199725 (B.L.), R01-DK091392 and R01-DK092461 (M.C.H. and O.W.M.), P30-DK07938 (O.W.M.), K99R00-DK094980 (C.H.), Cancer Prevention Research Institute of Texas grant RP120718 (B.L.) and a Fondation Leducq grant 15CBD04 (B.L., A.F.F. and S.S.).

**Reviewer information** Nature thanks D. Harrison and the other anonymous reviewer(s) for their contribution to the peer review of this work.

**Author contributions** A.F.F., S.S., Y.W., C.H., T.T., Y.L., M.C.H. and B.L. designed the study. A.F.F., S.S., Y.W., Z.Z., M.S., K.L.M. and W.-C.C. performed biochemical analyses. A.F.F., S.S. and Z.Z. performed autophagy microscopic analyses. A.F.F. and S.S. performed renal and cardiac histopathological analyses. G.B. characterized malignancies. A.F.F., S.S., D.K.M., G.G.S., G.B., O.W.M., M.C.H. and B.L. discussed and analysed data. A.F.F., S.S., M.C.H. and B.L. wrote the manuscript. A.F.F. and S.S. contributed equally and the order of these authors was determined arbitrarily.

**Competing interests** B.L. is a Scientific Founder of Casma Therapeutics, Inc.

## Additional information

**Extended data** is available for this paper at <https://doi.org/10.1038/s41586-018-0162-7>.

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41586-018-0162-7>.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>.

**Correspondence and requests for materials** should be addressed to M.C.H. or B.L. **Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## METHODS

**Cell culture.** HeLa cells were obtained from the ATCC (American Type Culture Collection), tested negative for mycoplasma by PCR and authenticated by ATCC Cell Line Authentication Service. Primary MEFs were isolated from mouse embryos at embryonic day (E) 13.5 and cultured as described<sup>29</sup>.

**Mice.** *Becn1*<sup>F121A/F121A</sup> knock-in mice were generated as described<sup>15</sup> and backcrossed for more than 12 generations to C57BL/6J mice (Jackson Laboratories). *Becn1*<sup>+/+</sup> (wild-type) and *Becn1*<sup>F121A/F121A</sup> (knock-in) littermate mice were crossed with GFP-LC3 transgenic animals<sup>18</sup> in a pure C57BL/6J background and tissues of offspring were used for autophagic flux analyses. Klotho hypomorphic mice (best known as *kl/kl*; referred to as *Kl*<sup>HM/HM</sup> in this manuscript) have been previously described<sup>3</sup> and were crossed with *Becn1*<sup>+/+</sup> (wild-type) and *Becn1*<sup>F121A/F121A</sup> (knock-in) mice to obtain double mutants. Mice were genotyped for the *Kl* and *Becn1* alleles as described<sup>3,15</sup>. All mice were housed on a 12-h light/dark cycle and both males and females were used for all analyses. For sample size and randomization information, please see the Life Sciences Reporting Summary. For survival analysis, mice were monitored weekly for the duration of the observation period. For western blot autophagy analyses, animals were starved overnight and re-fed for 3 h before sample collection to minimize variability due to differences in food intake. All animal procedures were performed in accordance with institutional guidelines and with approval from the UT Southwestern Medical Center Institutional Animal Care and Use Committee.

**Immunoprecipitations.** For beclin 1–BCL2 co-immunoprecipitation, frozen tissues were weighed and homogenized in ice-cold lysis buffer (25 mM HEPES, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100; 1 ml per 100 mg tissue) containing cOmplete, mini protease (Roche) and Halt phosphatase (Thermo Scientific) inhibitor cocktails for 30 min at 4 °C. Lysates were centrifuged (16,000g at 4 °C for 30 min) and the supernatants were pre-cleared with 60 µl protein-G agarose beads for 2 h and incubated overnight with BCL2-agarose (or IgG) antibody. Immunoprecipitates were washed five times, resuspended in 2× SDS–PAGE loading buffer, boiled for 5 min and analysed by western blot using anti-beclin 1 (sc-7382, Santa Cruz; 1:500 dilution), anti-BCL2 (sc-7382, Santa Cruz; 1:100 dilution), anti-klotho (KO603, Trans Genic Inc. Japan; 1:1,000 dilution) and anti-actin (sc-47778, Santa Cruz; 1:5,000 dilution) antibodies.

For in vitro analyses of the effects of klotho on beclin 1 and BCL2 binding, soluble full-length mouse klotho protein was purified as previously described<sup>27</sup>, and HeLa (ATCC) cells were treated either with PBS or klotho protein (0.4 nM or 2.0 nM) for 24 h before co-immunoprecipitation using the same protocol described above for mouse tissues.

**Autophagy analyses.** To assess in vivo autophagy levels in different tissues, two- and six-month-old homozygous *Becn1*<sup>F121A/F121A</sup>;GFP-LC3 or *Becn1*<sup>+/+</sup>;GFP-LC3 mice (three mice per group) were treated with either PBS or chloroquine (50 mg kg<sup>−1</sup>) for 6 h. Mice were then perfused with 4% paraformaldehyde (PFA) in PBS and tissues were collected and processed for frozen sectioning as described<sup>18</sup>. The total number of GFP–LC3 puncta was counted per 2,500 µm<sup>2</sup> area (more than 15 randomly chosen fields were used per mouse) and the average value for each tissue for each mouse was determined by an observer blinded to genotype. The mouse muscle, heart, and liver tissue sections were imaged using a 63× objective and for renal glomeruli and proximal convoluted tubules, tissue sections were imaged using a 40× objective on a Zeiss AxioPlan 2 microscope.

For western blot analysis, frozen tissues were lysed in ice-cold lysis buffer (Tris-HCl, pH 7.5, 50 mM, NaCl 150 mM, 1 mM EDTA, 1% Triton X-100) with cOmplete, mini protease (Roche) and Halt phosphatase (Thermo Scientific) inhibitor cocktails for 30 min at 4 °C. Lysates were centrifuged at 16,000g for 30 min. Cleared lysates were diluted in 2× SDS–PAGE loading buffer and analysed using anti-p62 (GP62-C, Progen, 1:1,000 dilution), anti-LC3B (L7543, Sigma, 1:10,000 dilution) and anti-actin (sc-47778, Santa Cruz, 1:5,000 dilution) antibodies. Endogenous LC3 immunofluorescence staining of paraffin-embedded tissues was performed as previously described<sup>30</sup>.

Electron microscopic analyses of MEFs derived from wild-type and knock-in littermate mice was performed and analysed as described<sup>16</sup>.

**Measurement of endocytosis.** Analysis of transferrin uptake by wild-type and knock-in MEFs as an indicator of endocytosis activity was performed as previously described<sup>31,32</sup>.

**Histopathological analyses.** Wild-type and knock-in control littermates aged two months and 20 months were perfused with 4% PFA in PBS before tissue collection,

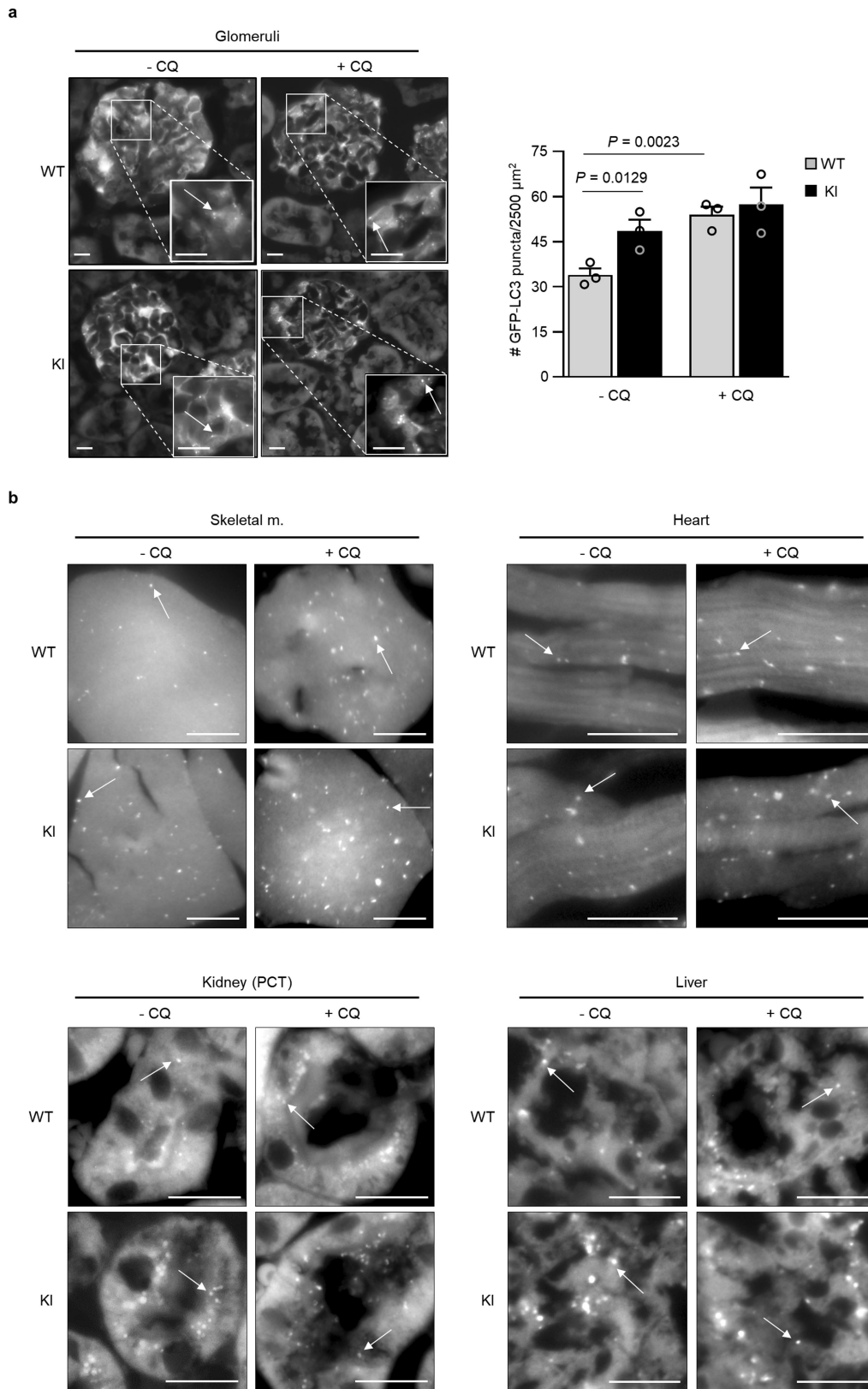
fixation, and preparation of paraffin-embedded sections for histopathological analyses. Standard haematoxylin and eosin (H&E) staining was performed for analyses of age-related neoplasia and renal histopathological score. H&E-stained tissues sections of several organs were evaluated per mouse in a genotype-blinded manner by a pathologist with expertise in the diagnosis of human and murine neoplasms and all cases of lymphoproliferative disease and non-lymphoid malignancies were recorded. TUNEL staining was performed according to manufacturer's instructions (ApopTag Peroxidase In situ Apoptosis Detection Kit, Millipore) and active caspase 3 staining was performed using anti-active caspase 3 (ab3202, Abcam, 1:20 dilution) antibody overnight at 4 °C and an ABC kit according to manufacturer's instructions. The total numbers of TUNEL-positive nuclei and active caspase 3-positive cells were counted in sections of the entire renal cortex of each mouse to calculate the number of TUNEL-positive nuclei and active caspase 3-positive cells per unit area, and the total numbers of TUNEL-positive nuclei and active caspase 3-positive cells were counted in each longitudinal section of the heart to calculate the number of TUNEL-positive nuclei and active caspase 3-positive cells per unit area. To determine the renal pathological score, ten random fields of the renal cortex sections were evaluated per mouse. Each field was given a pathological score using the following scale: 0, absence of damage; 1, <15% tissue area with damage; 2, 15–50% tissue area with damage; 3, >50% tissue area with damage. The scores of each field were summed to give a final histopathological score for each mouse, ranging from 0 to 30. Wheat germ agglutinin (WGA) staining was performed by pre-incubating slides in citrate buffer at 50 °C for 13 min, blocking with 1% bovine serum albumin (BSA), 5% goat serum in PBS for 1 h, and staining with Alexa Fluor 594-WGA (W11262, ThermoFisher Scientific, 1:100 dilution) for 1 h at room temperature. The average cross-sectional size of cardiomyocytes was determined by ImageJ; 100 cells were analysed per mouse (using 20 cells per field and acquiring 5 fields per heart). Masson's trichrome staining was performed according to the manufacturer's instructions (ab150686, Abcam) and used for analyses of renal cortex and cardiac fibrosis. The percentage area of tissue with fibrosis was measured. The fibrotic areas and total field size were manually outlined and quantified using ImageJ for 10 randomly chosen fields for the renal cortex per mouse and 5 randomly chosen fields for the heart per mouse. Quantification of all histopathological analyses was performed by an observer blinded to experimental age and genotype.

**Statistical analyses.** Data were analysed using the GraphPad Prism 7 and OASIS 2 software. log-rank (Mantel–Cox) tests were used to analyse Kaplan–Meier curves, and a Booshloo's test for maximum lifespan analysis (at 90% survival) was performed as described<sup>33</sup>. Chi-square test was used to compare the percentage of mice with spontaneous malignancies. One-tailed unpaired Student's *t*-tests were used for analyses of beclin 1–BCL2 binding and autophagy in wild-type versus knock-in mice. For data with a non-normal distribution, the ROUT (robust regression and outlier removal) method<sup>34</sup> was used to eliminate outliers with Q coefficient set at 1% and data were analysed by two-tailed Mann–Whitney tests.

**Reporting summary.** Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

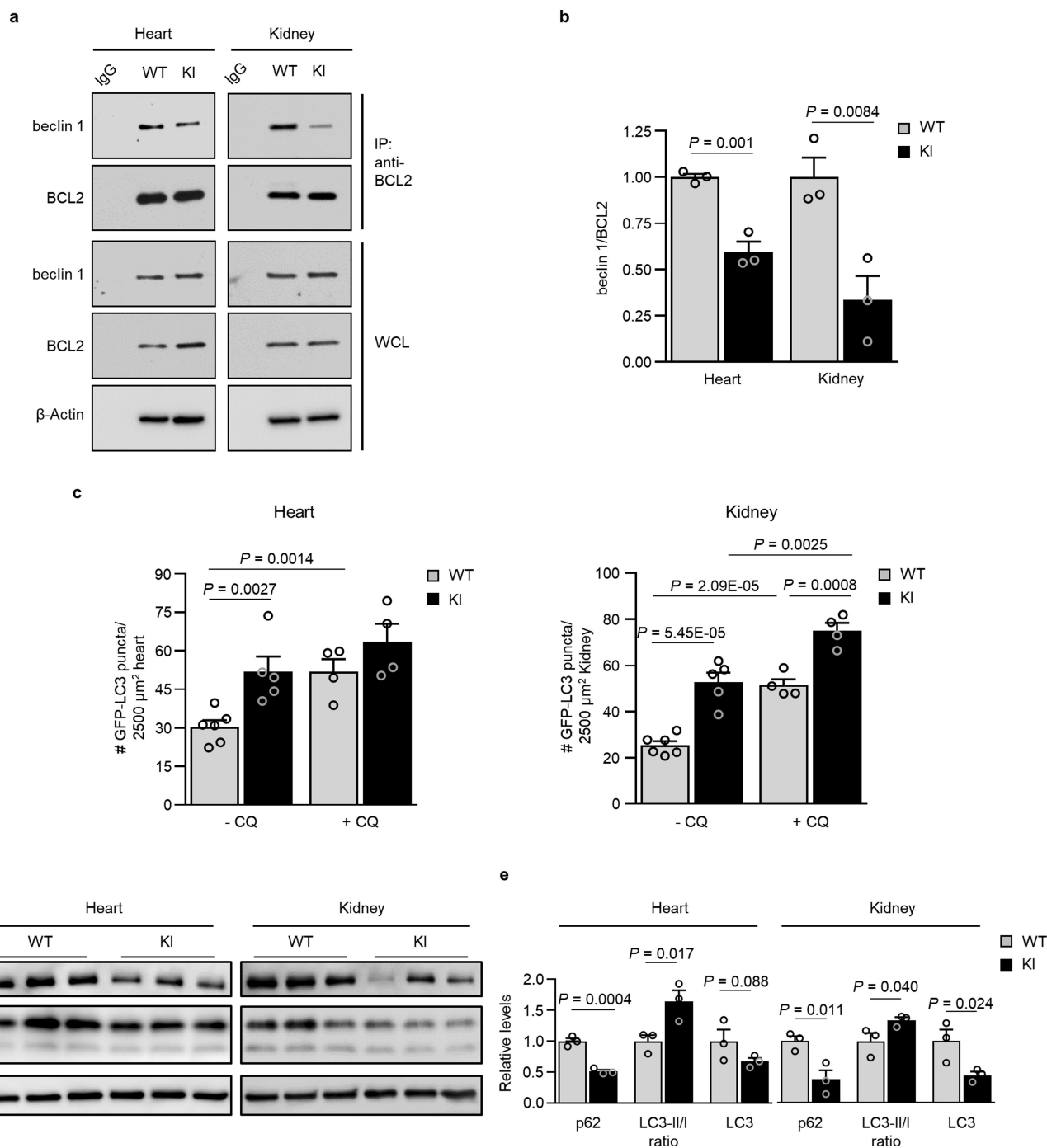
**Data availability.** Full scans for all western blots are provided in Supplementary Fig. 1. Source data for all graphs in this manuscript have been provided. All other data are available from the corresponding authors on reasonable request.

29. Su, T. et al. Deletion of histidine triad nucleotide-binding protein 1/PKC-interacting protein in mice enhances cell growth and carcinogenesis. *Proc. Natl Acad. Sci. USA* **100**, 7824–7829 (2003).
30. Sebt, S. et al. BAT3 modulates p300-dependent acetylation of p53 and autophagy-related protein 7 (ATG7) during autophagy. *Proc. Natl Acad. Sci. USA* **111**, 4115–4120 (2014).
31. Fielding, A. B., Willox, A. K., Okeke, E. & Royle, S. J. Clathrin-mediated endocytosis is inhibited during mitosis. *Proc. Natl Acad. Sci. USA* **109**, 6572–6577 (2012).
32. Tacheva-Grigorova, S. K., Santos, A. J., Boucrot, E. & Kirchhausen, T. Clathrin-mediated endocytosis persists during unperturbed mitosis. *Cell Reports* **4**, 659–668 (2013).
33. Wang, C., Li, Q., Redden, D. T., Weindruch, R. & Allison, D. B. Statistical methods for testing effects on “maximum lifespan”. *Mech. Ageing Dev.* **125**, 629–632 (2004).
34. Motulsky, H. J. & Brown, R. E. Detecting outliers when fitting data with nonlinear regression — a new method based on robust nonlinear regression and the false discovery rate. *BMC Bioinformatics* **7**, 123 (2006).



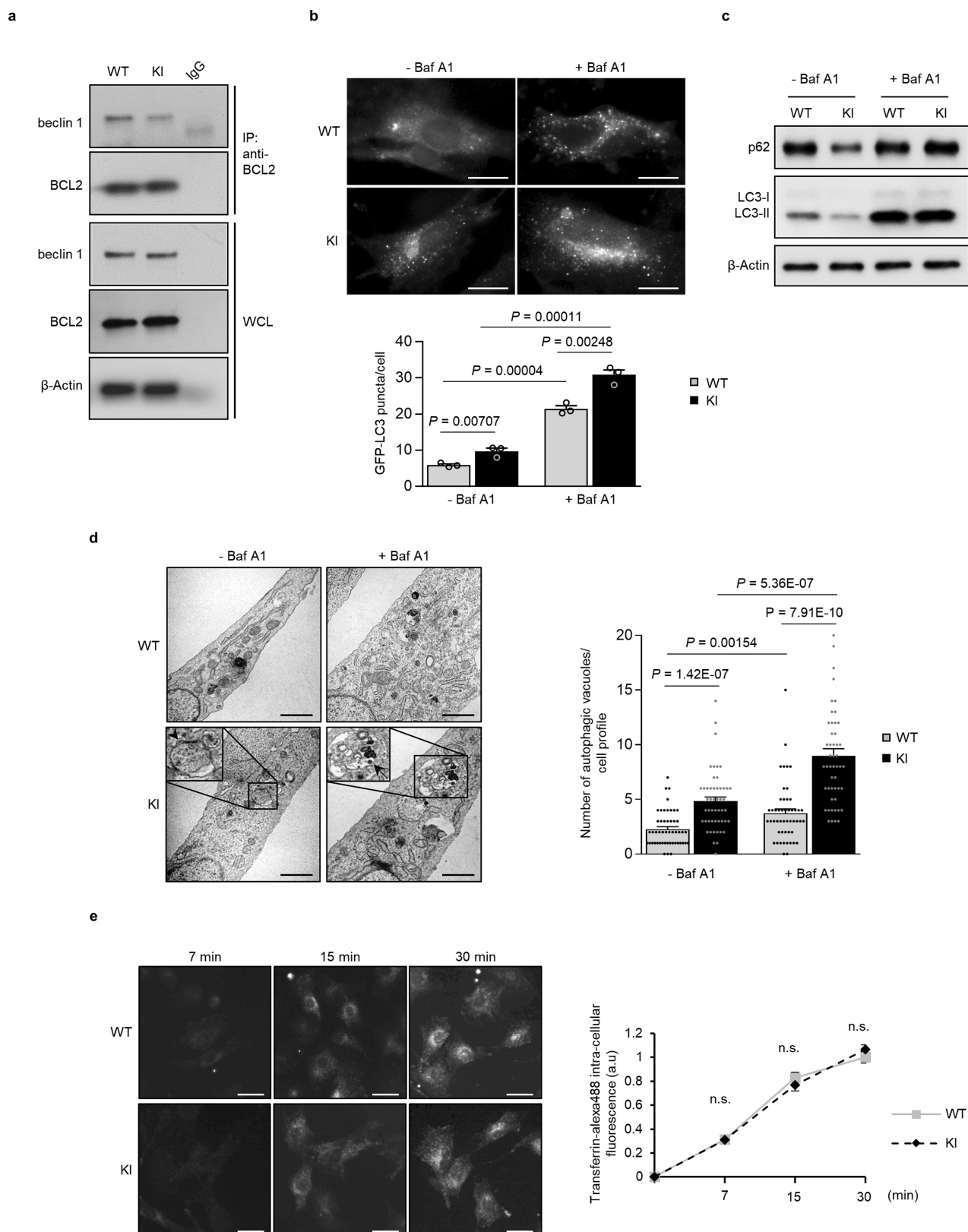
**Extended Data Fig. 1 | Increased basal autophagy in tissues of beclin 1(F121A) knock-in mice. a,** Representative images and quantification of GFP-LC3 puncta (autophagosomes) in glomeruli from *Becn1*<sup>+/+</sup>;GFP-LC3 (wild-type, WT) and *Becn1*<sup>F121A/F121A</sup>;GFP-LC3 (knock-in, KI) mice with

or without 50 mg kg<sup>-1</sup> chloroquine for 6 h. Scale bars, 10  $\mu\text{m}$ . Data are mean  $\pm$  s.e.m. for three mice per genotype. **b,** Enlarged versions of the images shown in Fig. 1c. *P* values determined by one-sided unpaired *t*-test. Arrows denote representative GFP-LC3 puncta.



**Extended Data Fig. 2 | Sustained increase in basal autophagy during adulthood in beclin 1(F121A) knock-in mice.** **a**, Co-immunoprecipitation of beclin 1 and BCL2 in representative samples of hearts and kidneys from eight-month-old *Becn1*<sup>+/+</sup> (wild-type) and *Becn1*<sup>F121A/F121A</sup> (knock-in) animals. **b**, Quantification of beclin 1 co-immunoprecipitated with BCL2 in indicated tissues of eight-month-old wild-type and knock-in mice ( $n = 3$  mice per genotype). **c**, Quantification of GFP-LC3 puncta in hearts and tissues from six-month-old

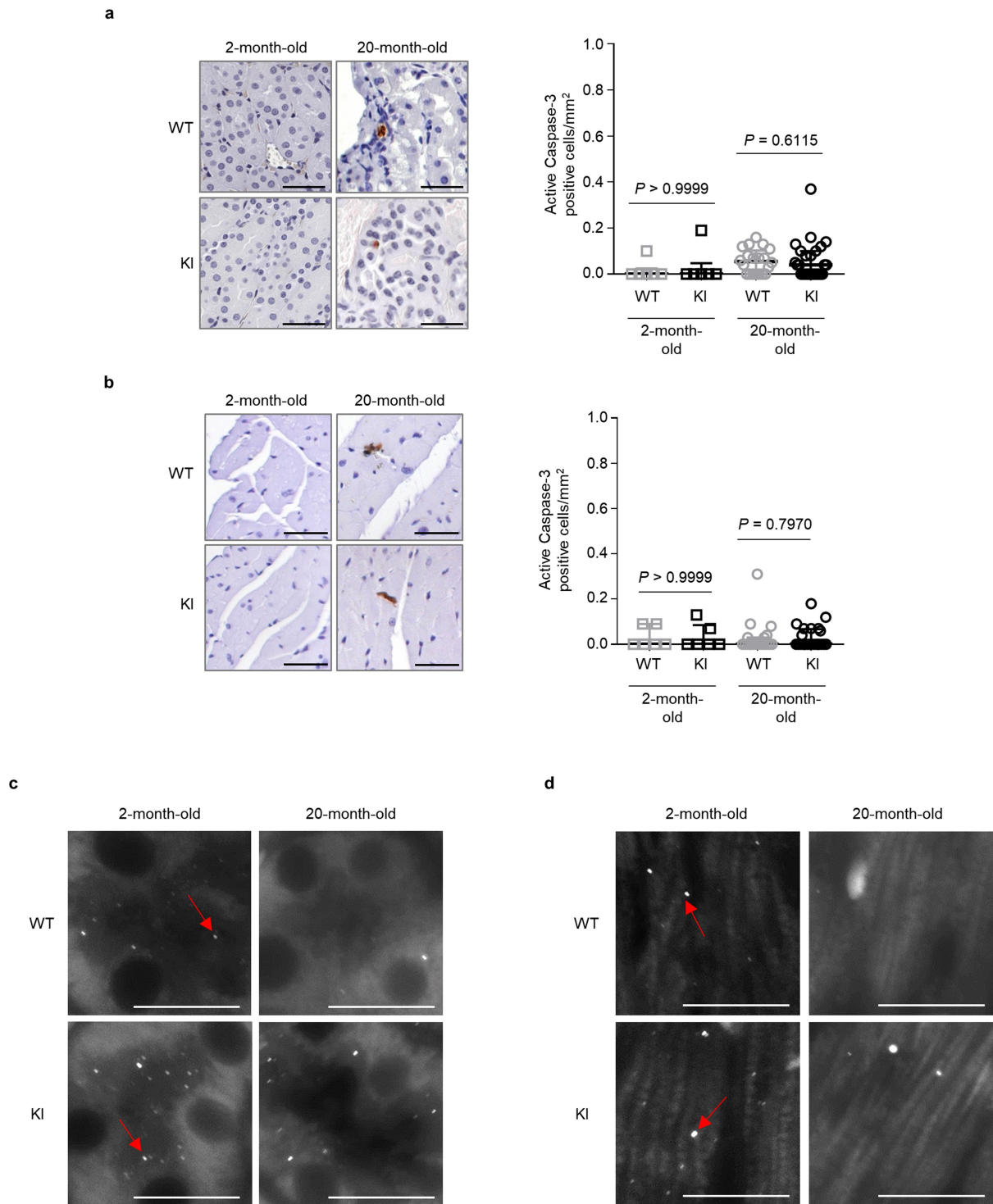
*Becn1*<sup>+/+</sup>;GFP-LC3 (wild-type) and *Becn1*<sup>F121A/F121A</sup>;GFP-LC3 (knock-in) mice with or without chloroquine (50 mg kg<sup>-1</sup>, 6 h). **d**, Western blot analysis of autophagy markers in the hearts and kidneys from eight-month-old wild-type and knock-in mice. Each lane represents a different mouse. **e**, Quantification of p62 and total LC3 levels (normalized to  $\beta$ -actin), as well as LC3-II/LC3-I ratios from samples in **d**. Data are mean  $\pm$  s.e.m. for three mice per genotype.  $P$  values were determined by one-sided unpaired  $t$ -test. For uncropped gels, see Supplementary Fig. 1.



Extended Data Fig. 3 | See next page for caption.

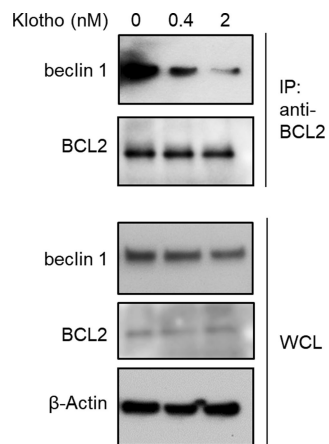
**Extended Data Fig. 3 | Increased autophagy, but not endocytosis, in beclin 1 (F121A) MEFs.** **a**, Co-immunoprecipitation of beclin 1 and BCL2 in MEFs derived from *Becn1*<sup>+/+</sup> (wild-type) and *Becn1*<sup>F121A/F121A</sup> (knock-in) animals. **b**, Representative images (top) and quantification (bottom) of GFP-LC3 puncta in wild-type and knock-in cells with or without 10 nM bafilomycin A1 (BafA1) for 3 h. Scale bars, 10  $\mu$ m. **c**, Western blot analysis of autophagy markers in wild-type and knock-in MEFs with or without BafA1 (100 nM, 2 h). **d**, Representative images and quantitative electron microscopic analysis of autophagic structures in wild-type and knock-in MEFs with or without BafA1 (100 nM, 3 h). Insets show representative autophagosome (arrowhead) and autolysosome

(arrow). Scale bars, 1  $\mu$ m. **e**, Representative images and quantification of transferrin uptake kinetics in wild-type and knock-in cells. Scale bars, 20  $\mu$ m. Results shown are representative of two and four independent experiments respectively for **a** and **c**. Data are mean  $\pm$  s.e.m. for three replicates in **b** and for 50 cells per genotype and condition in **d**. Data points on line graph in **e** denote mean  $\pm$  s.e.m. for cells at 7 min ( $n = 63$ , WT;  $n = 54$ , KI), 15 min ( $n = 58$ , WT;  $n = 52$  KI), and 30 min ( $n = 76$ , WT;  $n = 83$ , KI). *P* values determined by unpaired one-sided (**b**) and two-sided (**d**, **e**) *t*-test. For uncropped gels, see Supplementary Fig. 1.



**Extended Data Fig. 4 | Apoptosis and autophagy analyses in kidneys and hearts of aged mice.** **a, b,** Representative images and quantification of active caspase 3-positive cells in kidneys (**a**) and hearts (**b**) from *Becn1*<sup>+/+</sup> (wild-type) and *Becn1*<sup>F121A/F121A</sup> (knock-in) animals. Two month-old wild-type and knock-in mouse kidneys and hearts ( $n = 6$  per genotype) were analysed. For kidney analyses, aged (20-month-old) wild-type

( $n = 20$ ) and knock-in ( $n = 26$ ) mice were used. For heart analyses, aged (20-month-old) wild-type ( $n = 19$ ) and knock-in ( $n = 26$ ) mice were used. Scatter plot bars represent median  $\pm$  interquartile ranges.  $P$  values were determined by two-sided Mann-Whitney test. **c, d,** Enlarged versions of the endogenous LC3 puncta (autophagosome) images shown in Fig. 3d (**c**) and Fig. 3h (**d**). Arrows denote representative LC3 puncta.



**Extended Data Fig. 5 | In vitro klotho treatment disrupts the beclin 1–BCL2 interaction.** Co-immunoprecipitation of beclin 1 and BCL2 in HeLa cells treated with PBS or the indicated concentrations of recombinant full-length mouse klotho protein for 24 h. Result shown is representative of two independent experiments. For uncropped gels, see Supplementary Fig. 1.

**Extended Data Table 1 | Increased median lifespan and maximal lifespan in beclin 1(F121A) knock-in mice**

Sex	<i>Becn1</i> genotype	Median lifespan	Maximum lifespan	
		Months (95% C.I.)	Months	Boschloo's test
All	WT	26 (25.0 ~ 26.0)	36	P = 0.0341
	KI	29 (28.0 ~ 29.0)	39	
Females	WT	27 (26.0 ~ 28.0)	34	P = 0.0605
	KI	30 (28.0 ~ 30.0)	39	
Males	WT	25 (24.0 ~ 25.0)	36	P = 0.0881
	KI	28 (26.0 ~ 29.0)	39	

# Hippo/Mst signalling couples metabolic state and immune function of CD8 $\alpha^+$ dendritic cells

Xingrong Du<sup>1</sup>, Jing Wen<sup>1</sup>, Yanyan Wang<sup>1</sup>, Peer W. F. Karmaus<sup>1</sup>, Alireza Khatamian<sup>2</sup>, Haiyan Tan<sup>3,4</sup>, Yuxin Li<sup>3,4</sup>, Cliff Guy<sup>1</sup>, Thanh-Long M. Nguyen<sup>1</sup>, Yogesh Dhungana<sup>1</sup>, Geoffrey Neale<sup>5</sup>, Junmin Peng<sup>3,4</sup>, Jiyang Yu<sup>2\*</sup> & Hongbo Chi<sup>1\*</sup>

**Dendritic cells orchestrate the crosstalk between innate and adaptive immunity. CD8 $\alpha^+$  dendritic cells present antigens to CD8 $^+$  T cells and elicit cytotoxic T cell responses to viruses, bacteria and tumours<sup>1</sup>. Although lineage-specific transcriptional regulators of CD8 $\alpha^+$  dendritic cell development have been identified<sup>2</sup>, the molecular pathways that selectively orchestrate CD8 $\alpha^+$  dendritic cell function remain elusive. Moreover, metabolic reprogramming is important for dendritic cell development and activation<sup>3,4</sup>, but metabolic dependence and regulation of dendritic cell subsets are largely uncharacterized. Here we use a data-driven systems biology algorithm (NetBID) to identify a role of the Hippo pathway kinases Mst1 and Mst2 (Mst1/2) in selectively programming CD8 $\alpha^+$  dendritic cell function and metabolism. Our NetBID analysis reveals a marked enrichment of the activities of Hippo pathway kinases in CD8 $\alpha^+$  dendritic cells relative to CD8 $\alpha^-$  dendritic cells. Dendritic cell-specific deletion of Mst1/2—but not Lats1 and Lats2 (Lats1/2) or Yap and Taz (Yap/Taz), which mediate canonical Hippo signalling—disrupts homeostasis and function of CD8 $^+$  T cells and anti-tumour immunity. Mst1/2-deficient CD8 $\alpha^+$  dendritic cells are impaired in presentation of extracellular proteins and cognate peptides to prime CD8 $^+$  T cells, while CD8 $\alpha^-$  dendritic cells that lack Mst1/2 have largely normal function. Mechanistically, compared to CD8 $\alpha^-$  dendritic cells, CD8 $\alpha^+$  dendritic cells exhibit much stronger oxidative metabolism and critically depend on Mst1/2 signalling to maintain bioenergetic activities and mitochondrial dynamics for their functional capacities. Further, selective expression of IL-12 by CD8 $\alpha^+$  dendritic cells depends on Mst1/2 and the crosstalk with non-canonical NF- $\kappa$ B signalling. Our findings identify Mst1/2 as selective drivers of CD8 $\alpha^+$  dendritic cell function by integrating metabolic activity and cytokine signalling, and highlight that the interplay between immune signalling and metabolic reprogramming underlies the unique functions of dendritic cell subsets.**

CD8 $\alpha^+$  dendritic cells (DCs) exhibit a preference for priming CD8 $^+$  T cells over CD4 $^+$  T cells, whereas CD8 $\alpha^-$  DCs are more efficient in priming CD4 $^+$  T cells<sup>5</sup>. To identify subset-specific regulators of DCs, we developed a systems biology tool called data-driven network-based Bayesian inference of drivers (NetBID), by integrating data from transcriptomics, whole proteomics and phosphoproteomics (Fig. 1a). Specifically, we computationally reconstructed a DC-specific signalling interactome (DCI) from a collective cohort of transcriptomic profiles of total DCs (Extended Data Fig. 1a) using information theory-based approaches<sup>6,7</sup>. Next, we superimposed the DCI with the transcriptome, proteome and phosphoproteome of CD8 $\alpha^+$  and CD8 $\alpha^-$  DCs. We hypothesized that if a signalling protein is a unique driver between DC subsets, its regulons in the DCI should be enriched in the differentially expressed genes and proteins, although the driver itself is not necessarily differentially expressed. Given their crucial roles in immune function<sup>8</sup>, we focused on protein kinases and identified 36 hub kinases whose regulons in the DCI were enriched in CD8 $\alpha^+$  versus CD8 $\alpha^-$  DC signatures in all of the transcriptome, proteome and phosphoproteome

profiles (Extended Data Fig. 1b, c). There was a striking enrichment of Hippo signalling<sup>9</sup> (Extended Data Fig. 1b, d), as many kinases involved in Hippo signalling (Extended Data Fig. 1e) were identified by NetBID, including Mst1 (also known as Stk4). Immunoblot analysis showed that CD8 $\alpha^+$  DCs had increased phosphorylation of Mst1/2 and Yap, and increased expression of Lats1 in comparison to CD8 $\alpha^-$  DCs (Fig. 1b). Moreover, the predicted regulons of Mst1 (Extended Data Fig. 1f) were considerably dysregulated upon deletion of Mst1/2 in total, CD8 $\alpha^+$  and CD8 $\alpha^-$  DCs (Fig. 1c and Extended Data Fig. 1g, h). Using this unbiased approach to capture putative master regulators, we have identified the marked enrichment of Hippo signalling in CD8 $\alpha^+$  DCs.

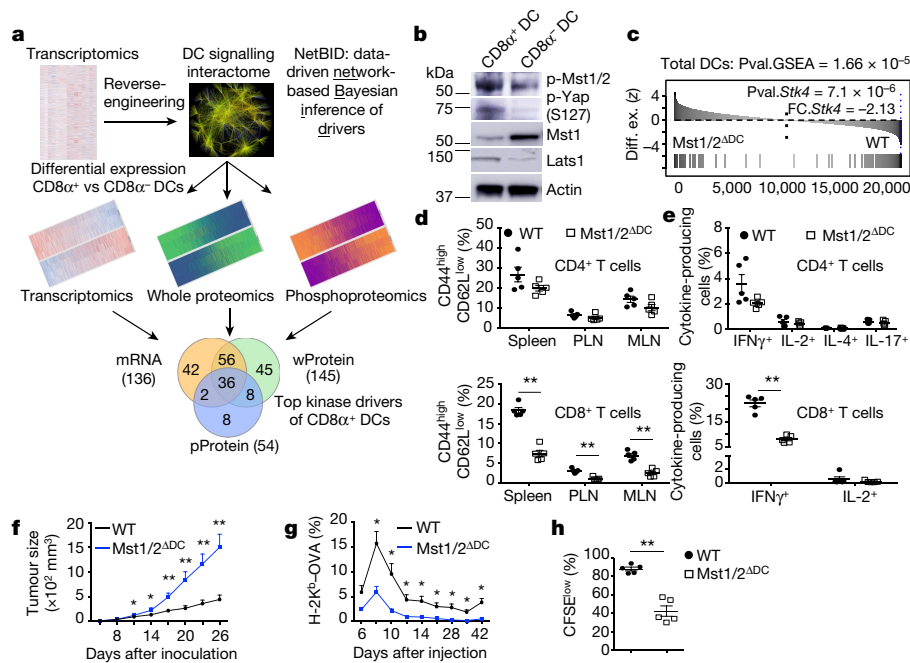
To systemically dissect the Hippo pathway in DCs, we engineered DC-specific deletion of Mst1/2, Lats1/2 or Yap/Taz using CD11c-Cre mice, resulting in Mst1/2 $^{\Delta DC}$ , Lats1/2 $^{\Delta DC}$ , or Yap/Taz $^{\Delta DC}$  mice, respectively. The cellularities of lymphoid organs and T cells were reduced in Mst1/2 $^{\Delta DC}$  mice (Extended Data Fig. 2a, b), but not in Lats1/2 $^{\Delta DC}$  and Yap/Taz $^{\Delta DC}$  mice (data not shown). Mst1/2 $^{\Delta DC}$  mice showed a decreased percentage of CD8 $^+$  T cells (Extended Data Fig. 2c), which was associated with a reduced effector/memory T cell population (Fig. 1d, Extended Data Fig. 2d), but CD4 $^+$  T cell homeostasis was largely unperturbed. Furthermore, CD44<sup>high</sup>CD8 $^+$  T cells from Mst1/2 $^{\Delta DC}$  mice expressed less IFN $\gamma$ , while CD44<sup>high</sup>CD4 $^+$  T cells had slightly reduced IFN $\gamma$  but normal IL-2, IL-4 and IL-17A expression (Fig. 1e, Extended Data Fig. 2e). In contrast to Mst1/2 $^{\Delta DC}$  mice, Lats1/2 $^{\Delta DC}$  and Yap/Taz $^{\Delta DC}$  mice exhibited normal T cell homeostasis (Extended Data Fig. 2f–k). Also, deletion of Mst1 or Mst2 alone did not affect immune homeostasis (Extended Data Fig. 2l–n). We verified the specific loss of Mst1/2 expression in DCs from Mst1/2 $^{\Delta DC}$  mice (Extended Data Fig. 3a, b) and rescue of CD8 $^+$  T cell phenotypes in mixed bone marrow chimaeras comprised of wild-type and Mst1/2 $^{\Delta DC}$  bone marrow-derived cells (Extended Data Fig. 3c, d). These results show that DCs require Mst1/2 to selectively orchestrate CD8 $^+$  T cell homeostasis, and this occurs independently of the classical Hippo pathway.

After challenge with MC38 colon adenocarcinoma cells, Mst1/2 $^{\Delta DC}$  mice exhibited drastically increased tumour growth (Fig. 1f) and impaired IFN $\gamma$  expression in CD8 $^+$  T cells (Extended Data Fig. 4a, b), although expression of PD-1, LAG3 and TIM3 were unaltered (Extended Data Fig. 4c, d). Additionally, Mst1/2 $^{\Delta DC}$  mice infected with ovalbumin-expressing *Listeria monocytogenes* (LM-OVA) showed reduced CD8 $^+$  T cell responses (Fig. 1g and Extended Data Fig. 4e, f). Furthermore, following adoptive transfer of OVA-reactive CD8 $^+$  T cells (OT-I) and immunization with OVA, proliferation of OT-I cells was greatly impaired in Mst1/2 $^{\Delta DC}$  mice (Fig. 1h and Extended Data Fig. 4g). Together, these data show that Mst1/2 signalling in DCs is required to orchestrate CD8 $^+$  T cell-mediated immune responses in vivo.

Mst1/2 $^{\Delta DC}$  mice contained normal percentages of splenic conventional DCs (cDCs) and plasmacytoid DCs (pDCs) (Extended Data Fig. 5a). Within cDCs, the percentage and number of CD8 $\alpha^+$

<sup>1</sup>Department of Immunology, St Jude Children's Research Hospital, Memphis, TN, USA. <sup>2</sup>Department of Computational Biology, St Jude Children's Research Hospital, Memphis, TN, USA.

<sup>3</sup>Departments of Structural Biology and Developmental Neurobiology, St Jude Children's Research Hospital, Memphis, TN, USA. <sup>4</sup>St Jude Proteomics Facility, St Jude Children's Research Hospital, Memphis, TN, USA. <sup>5</sup>Hartwell Center for Bioinformatics and Biotechnology, St Jude Children's Research Hospital, Memphis, TN, USA. \*e-mail: [jiyang.yu@stjude.org](mailto:jiyang.yu@stjude.org); [hongbo.chi@stjude.org](mailto:hongbo.chi@stjude.org)



**Fig. 1 | NetBID identifies Hippo signalling kinases as drivers of CD8 $\alpha^+$  DCs, and deletion of Mst1/2 in DCs leads to selective CD8 $^+$  T cell homeostatic and functional defects. a, Overview of NetBID. b, Immunoblot of splenic CD8 $\alpha^+$  and CD8 $\alpha^-$  DCs. c, Enrichment of predicted Mst1 signalling regulons in differentially expressed genes between Mst1/2-deficient (Mst1/2 $^{\Delta DC}$ ) and wild-type (WT) DCs. Pval, *Stk4* and FC, *Stk4* indicate the P value and signed fold change of *Stk4* (also known as *Mst1*) expression. Diff. ex., differential expression. d, Frequencies of CD44 $^{high}$ CD62L $^{low}$  effector/memory cells in T cells**

from spleen, peripheral lymph nodes (PLN) and mesenteric lymph nodes (MLN) ( $n = 5$  per genotype). e, Frequencies of cytokine-producing cells ( $n = 5$  per genotype). f, MC38 tumour growth ( $n = 10$  for wild type,  $n = 6$  for Mst1/2 $^{\Delta DC}$ ). g, Frequency of blood H-2K $^b$ -OVA $^+$ CD8 $^+$  T cells from LM-OVA-infected mice ( $n = 5$  for wild type,  $n = 4$  for Mst1/2 $^{\Delta DC}$ ). h, Frequency of CFSE $^{low}$  proliferated cells of donor OT-I T cells in OVA-immunized mice ( $n = 5$  per genotype). Data are shown as mean and s.e.m. \* $P < 0.05$ , \*\* $P < 0.01$ ; two-tailed unpaired Student's *t*-test in d–h. Data summarize two (f–h), three (b) or four (d, e) independent experiments.

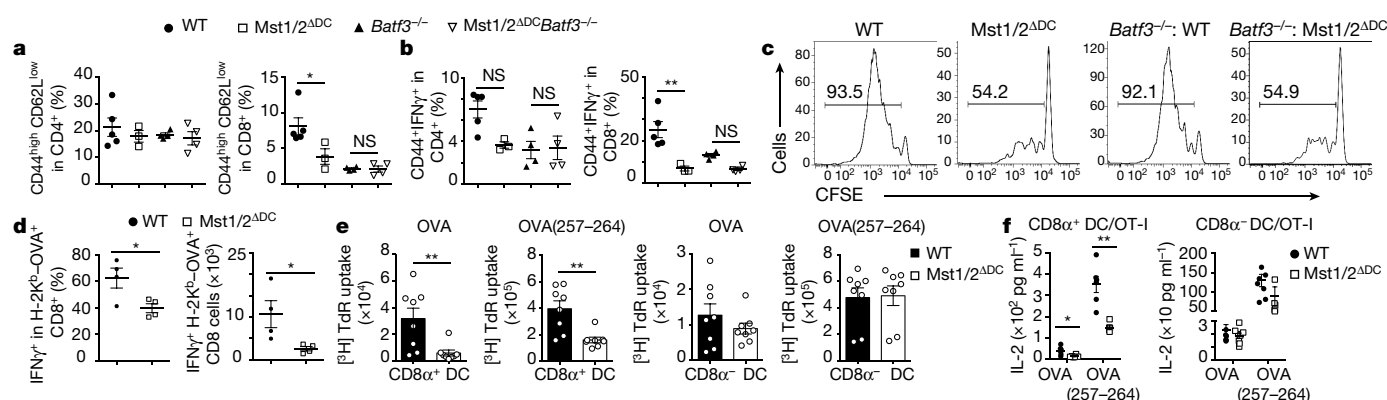
DCs increased and those of CD8 $\alpha^-$  DCs decreased (Extended Data Fig. 5b). Mst1/2-deficient DCs showed normal expression of multiple surface molecules, except for a slight reduction of PD-L1 on CD8 $\alpha^-$  DCs (Extended Data Fig. 5c). Moreover, mixed bone marrow chimera experiments revealed a cell-intrinsic role of Mst1/2 in DC homeostasis (Extended Data Fig. 5d–f). However, loss of Mst1 or Mst2 alone, or deletion of Lats1/2 or Yap/Taz, did not affect DC homeostasis (Extended Data Fig. 6a–c). Therefore, Mst1/2 control CD8 $\alpha^+$  DC homeostasis independently of the classical Hippo pathway.

We hypothesized that a selective functional defect of CD8 $\alpha^+$  DCs in Mst1/2 $^{\Delta DC}$  mice accounts for the altered CD8 $^+$  T cell homeostasis and activation. The T cell compartment in mice lacking *Batf3*, whose deletion selectively ablates CD8 $\alpha^+$  DCs $^{10}$ , largely phenocopied those in Mst1/2 $^{\Delta DC}$  mice, with reduced effector/memory T cell population and reduced IFN $\gamma$  expression in CD8 $^+$  T cells (Fig. 2a, b). Additionally, these parameters were comparable between Mst1/2-sufficient and Mst1/2-deficient mice in the *Batf3* $^{-/-}$  background (Fig. 2a, b). We generated *Batf3* $^{-/-}$ :Mst1/2 $^{\Delta DC}$  mixed bone marrow chimaeras, following an established strategy $^{11}$ , to restrict Mst1/2 deficiency to CD8 $\alpha^+$  DCs only, and found that the chimaeras were impaired in supporting OT-I CD8 $^+$  T-cell priming to the same extent as Mst1/2 $^{\Delta DC}$  complete chimaeras, indicating a selective defect in CD8 $\alpha^+$  DCs lacking Mst1/2 (Fig. 2c and Extended Data Fig. 7a). To further test this notion, we transferred OT-I T cells into mice deficient in  $\beta_2$ -microglobulin and thus MHC-I, and immunized these mice with OVA-pulsed CD8 $\alpha^+$  or CD8 $\alpha^-$  DCs. Mst1/2-deficient CD8 $\alpha^+$  DCs induced significantly weaker proliferation of OT-I T cells than wild-type counterparts, whereas wild-type and Mst1/2-deficient CD8 $\alpha^-$  DCs had comparable function (Extended Data Fig. 7b). Finally, in an in vivo cross-presentation assay, IFN $\gamma$  production from endogenous antigen-specific CD8 $^+$  T cells was significantly reduced in Mst1/2 $^{\Delta DC}$  mice (Fig. 2d). Of note, Mst1/2-deficient DCs had no defects in antigen uptake or cell survival (Extended Data Fig. 7c–f).

Therefore, Mst1/2 are required for in vivo cross-presentation ability of CD8 $\alpha^+$  DCs.

In vitro, Mst1/2-deficient CD8 $\alpha^+$  DCs showed a profound defect in mediating OT-I T cell proliferation in response to OVA protein, and to a lesser extent, OVA(257–264) peptide (Fig. 2e), whereas wild-type and Mst1/2-deficient CD8 $\alpha^-$  DCs had comparable function (Fig. 2e). Similar phenotypes were observed for IL-2 secretion from OT-I T cells (Fig. 2f). By contrast, either CD8 $\alpha^+$  or CD8 $\alpha^-$  DCs from Mst1/2 $^{\Delta DC}$  mice were equivalent to wild-type counterparts in priming proliferation of OVA-specific CD4 $^+$  (OT-II) T cells (Extended Data Fig. 7g). Mst1/2 deficiency also impaired the CD8 $^+$  T cell-priming function of FLT3L-cultured CD24 $^{high}$  bone marrow-derived dendritic cells (BMDCs), which are functionally equivalent to splenic CD8 $\alpha^+$  DCs $^{12}$ , but not that of CD24 $^{low}$  cells or GM-CSF-derived BMDCs (Extended Data Fig. 7h–j). Moreover, CD8 $\alpha^+$  DCs treated with a specific Mst1/2 inhibitor were defective in priming of CD8 $^+$  T cells (Extended Data Fig. 7k). Therefore, Mst1/2 selectively program CD8 $\alpha^+$  DC functions to prime CD8 $^+$  T cells. However, Lats1/2 or Yap/Taz deficiency did not affect the T cell-priming function of CD8 $\alpha^+$  DCs or their responsiveness to Mst1/2 inhibition (Extended Data Fig. 7l, m), further establishing a role of Mst1/2 in non-canonical Hippo signalling.

Metabolic reprogramming is associated with DC development and activation $^{3,4}$ , but the metabolic requirements of different DC subsets are poorly defined. Our proteomics profiling revealed a significant enrichment of metabolic pathways in CD8 $\alpha^+$  DCs (Extended Data Fig. 8a). Consistent with these results, CD8 $\alpha^+$  DCs showed much higher oxygen consumption rate (OCR), extracellular acidification rate (ECAR), mitochondrial mass and membrane potential than CD8 $\alpha^-$  DCs (Fig. 3a and Extended Data Fig. 8b, c). We hypothesized that the unique metabolic state of CD8 $\alpha^+$  DCs contributes to their immune priming functions. Indeed, treatment of CD8 $\alpha^+$  DCs with metformin (an inhibitor of mitochondrial complex I activity) or 2-deoxyglucose (2-DG, an inhibitor of hexokinase activity) markedly impaired their



**Fig. 2 | Mst1/2 are selectively required in CD8 $\alpha^+$  DCs to orchestrate CD8 $^+$  T cell homeostasis and function.** **a**, **b**, Frequencies of CD44<sup>high</sup>CD62L<sup>low</sup> (**a**) and CD44<sup>+</sup>IFN $\gamma^+$  cells (**b**) in splenic T cells from wild-type ( $n=5$ ), Mst1/2<sup>ΔDC</sup> ( $n=3$ ), Batf3<sup>-/-</sup> ( $n=4$ ) and Mst1/2<sup>ΔDC</sup>Batf3<sup>-/-</sup> ( $n=4$ ) mice. **c**, CFSE dilution of donor OT-I T cells in wild-type, Mst1/2<sup>ΔDC</sup>, Batf3<sup>-/-</sup>:WT or Batf3<sup>-/-</sup>:Mst1/2<sup>ΔDC</sup> chimaeras immunized with OVA. **d**, Frequency and number of IFN $\gamma^+$  cells among H-2K<sup>b</sup>-OVA<sup>+</sup>CD8 $^+$  T cells after immunization with OVA-

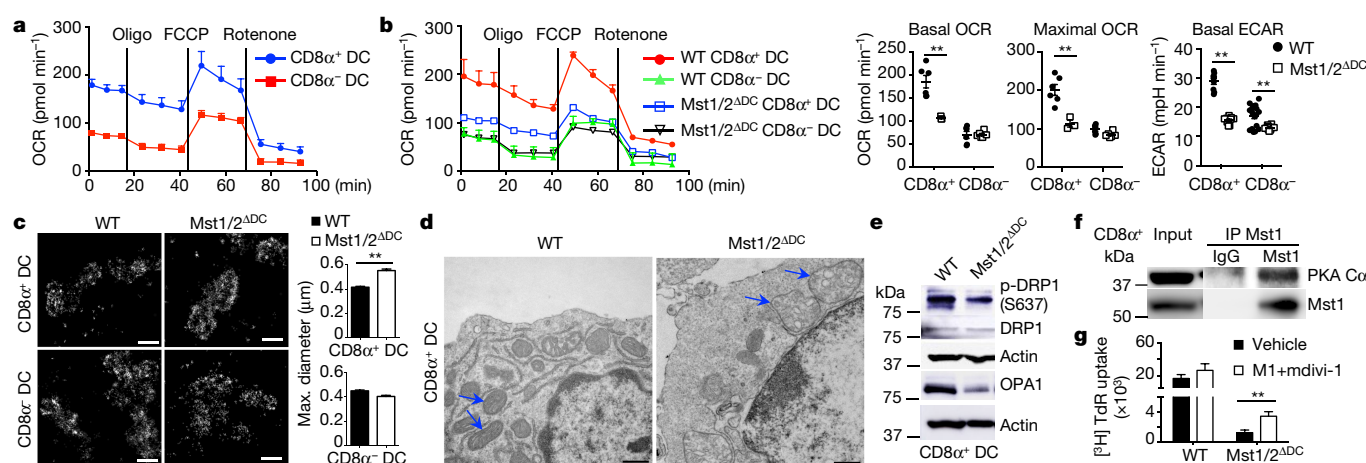
loaded irradiated B2m<sup>-/-</sup> splenocytes ( $n=4$  per genotype). **e**, Thymidine incorporation ([<sup>3</sup>H] TdR assay) in OT-I T cells cultured with CD8 $\alpha^+$  or CD8 $\alpha^-$  DCs pulsed with OVA protein or OVA(257–264) peptide ( $n=8$  per genotype). **f**, IL-2 from co-cultures in **e** ( $n=6$  per genotype for CD8 $\alpha^+$  DCs, and  $n=8$  per genotype for CD8 $\alpha^-$  DCs). Data are shown as mean and s.e.m. NS, not significant; \* $P<0.05$ , \*\* $P<0.01$ ; one-way ANOVA in **a**, **b**; two-tailed unpaired Student's  $t$ -test in **d–f**. Data summarize two (**c**, **f**), three (**d**, **e**) or four (**a**, **b**) independent experiments.

ability to prime OT-I T cells, whereas these inhibitors had only modest effects on CD8 $\alpha^-$  DCs (Extended Data Fig. 8d). Also, deficiency of the metabolic regulator mTOR dampened the CD8 $^+$  T cell-priming function of CD8 $\alpha^+$  DCs (Extended Data Fig. 8e). Therefore, CD8 $\alpha^+$  DCs have elevated metabolic activities relative to CD8 $\alpha^-$  DCs, which contribute to their functional capacity.

Furthermore, OCR and ECAR were drastically and selectively reduced in Mst1/2-deficient CD8 $\alpha^+$  DCs (Fig. 3b). Since the mitochondrion is the main organelle for oxidative phosphorylation, we measured the structural integrity of mitochondria. Mst1/2-deficient CD8 $\alpha^+$  DCs had enlarged mitochondria (Fig. 3c) and abnormal mitochondrial mass and membrane potential (Extended Data Fig. 8f). Transmission electron microscopy showed that Mst1/2-deficient CD8 $\alpha^+$  DCs exhibited increased mitochondrial size, but the mitochondrial cristae, which are required for efficient oxidative phosphorylation<sup>13,14</sup>, were markedly disorganized (Fig. 3d). By contrast, mitochondrial homeostasis

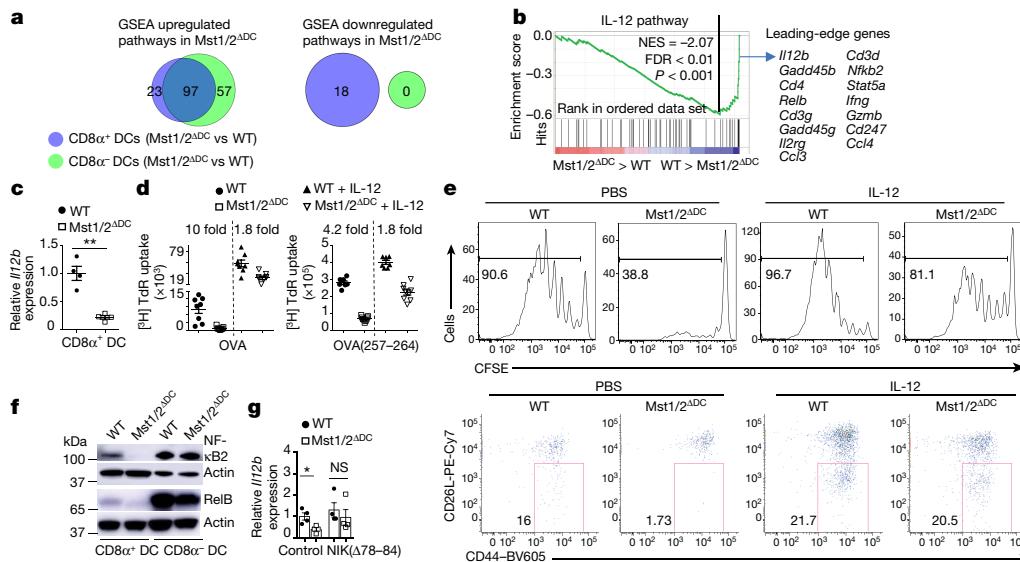
and structure were less affected by Mst1/2 deficiency in CD8 $\alpha^-$  DCs (Extended Data Fig. 8f, g). Consistent with the role of mitochondrial cristae in the assembly and stability of respiratory chain complexes<sup>13</sup>, NDUFB8 and MT-CO1, components of respiratory chain complexes I and IV respectively, showed impaired expression in Mst1/2-deficient CD8 $\alpha^+$  DCs (Extended Data Fig. 8h). Collectively, Mst1 and Mst2 orchestrate the metabolic activities and mitochondrial integrity of CD8 $\alpha^+$  DCs.

We investigated the molecular basis underlying mitochondrial functions in CD8 $\alpha^+$  DCs. We noted that Lats1/2-deficient DCs had normal mitochondrial profiles (Extended Data Fig. 8i). Additionally, mTORC1 activity and c-Myc expression were unaltered by Mst1/2 deficiency (Extended Data Fig. 8j), and expression of respiratory chain proteins was normal in mTOR-deficient DCs (Extended Data Fig. 8k). Given the regulation of mitochondrial morphology by fission and fusion processes (mitochondrial dynamics) and the link to oxidative



**Fig. 3 | Unique metabolic state of CD8 $\alpha^+$  DCs supports their immune priming function in an Mst1/2-dependent manner.** **a**, OCR of splenic CD8 $\alpha^+$  and CD8 $\alpha^-$  DCs. Oligo, oligomycin; FCCP, carbonyl cyanide  $p$ -trifluoromethoxyphenylhydrazone. **b**, OCR bioenergetics profile, basal and maximal OCR ( $n=6$  for wild-type CD8 $\alpha^+$  DCs,  $n=3$  for Mst1/2-deficient CD8 $\alpha^+$  DCs,  $n=6$  per genotype for CD8 $\alpha^-$  DCs), and basal ECAR ( $n=9$  per genotype for CD8 $\alpha^+$  DCs,  $n=18$  for wild-type CD8 $\alpha^-$  DCs,  $n=12$  for Mst1/2-deficient CD8 $\alpha^-$  DCs). **c**, Stochastic optical reconstruction microscopy (STORM) analysis of mitochondrial marker TOM20. Right, quantified maximal mitochondrial diameter. **d**, Transmission electron microscopy analysis of mitochondria (arrows) in CD8 $\alpha^+$  DCs. **e**, Immunoblot of CD8 $\alpha^+$  DCs. **f**, CD8 $\alpha^+$  DC lysate was immunoprecipitated with Mst1 antibody and blotted with PKA C $\alpha$ /C $\beta$  antibody. **g**, Thymidine incorporation in OT-I T cells cultured with OVA protein-pulsed wild-type ( $n=13$  from four mice) or Mst1/2<sup>ΔDC</sup> ( $n=11$  from four mice) CD8 $\alpha^+$  DCs pre-treated with vehicle or M1 + mdivi-1. Scale bars in **c** and **d**, 500 nm. Data are shown as mean and s.e.m. \*\* $P<0.01$ ; two-tailed unpaired Student's  $t$ -test in **b**, **c**, **g**. Data summarize two (**b**, **c**, **e–g**) or four (**a**) independent experiments.

**d**, Transmission electron microscopy analysis of mitochondria (arrows) in CD8 $\alpha^+$  DCs. **e**, Immunoblot of CD8 $\alpha^+$  DCs. **f**, CD8 $\alpha^+$  DC lysate was immunoprecipitated with Mst1 antibody and blotted with PKA C $\alpha$ /C $\beta$  antibody. **g**, Thymidine incorporation in OT-I T cells cultured with OVA protein-pulsed wild-type ( $n=13$  from four mice) or Mst1/2<sup>ΔDC</sup> ( $n=11$  from four mice) CD8 $\alpha^+$  DCs pre-treated with vehicle or M1 + mdivi-1. Scale bars in **c** and **d**, 500 nm. Data are shown as mean and s.e.m. \*\* $P<0.01$ ; two-tailed unpaired Student's  $t$ -test in **b**, **c**, **g**. Data summarize two (**b**, **c**, **e–g**) or four (**a**) independent experiments.



**Fig. 4 | Mst1/2 orchestrate selective expression of IL-12 in CD8 $\alpha^+$  DCs via crosstalk with non-canonical NF- $\kappa$ B signalling.** **a**, Venn diagram showing the overlap of significantly upregulated or downregulated pathways by GSEA analysis in CD8 $\alpha^+$  and CD8 $\alpha^-$  DCs. **b**, Underrepresentation of the IL-12 pathway in Mst1/2-deficient CD8 $\alpha^+$  DCs. NES, normalized enrichment score. **c**, *Il12b* expression in CD8 $\alpha^+$  DCs from wild-type and Mst1/2 $\Delta$ DC mice ( $n = 4$  per genotype). **d**, Thymidine incorporation of OT-I T cells cultured with CD8 $\alpha^+$  DCs pulsed with OVA protein or OVA(257–264) peptide with or without IL-12 (left,  $n = 8$  from three wild-type mice,  $n = 9$  from three Mst1/2 $\Delta$ DC mice; right,  $n = 7$  from

three wild-type mice,  $n = 9$  from three Mst1/2 $\Delta$ DC mice). The numbers above the graph are the relative ratios of wild-type versus Mst1/2 $\Delta$ DC DC groups. **e**, CFSE dilution (top) and CD62L/CD44 expression (bottom) of transferred OT-I T cells in OVA-immunized mice. **f**, NF- $\kappa$ B2 (p100) and RelB expression in DCs. **g**, Relative *Il12b* expression in wild-type or Mst1/2-deficient CD24<sup>high</sup> FLT3L-derived BMDCs transduced with control or NIK( $\Delta$ 78–84) virus ( $n = 4$  mice per group). Data are shown as mean and s.e.m. NS, not significant; \* $P < 0.05$ , \*\* $P < 0.01$ ; two-tailed unpaired Student's *t*-test in **c**, **g**. Data summarize two (**c–f**) or three (**g**) independent experiments.

phosphorylation<sup>15</sup>, we examined the role of Mst1/2 in mitochondrial dynamics. Mst1/2-deficient CD8 $\alpha^+$  DCs exhibited reduced expression of the fusion regulator OPA1<sup>15</sup>, and reduced phosphorylation of the fission protein DRP1 at Ser637, which is mediated by protein kinase A (PKA) to inhibit mitochondrial fission<sup>15</sup> (Fig. 3e). Moreover, Mst1 interacted with PKA in CD8 $\alpha^+$  DCs (Fig. 3f). Thus, in line with the disorganized cristae, Mst1/2-deficient CD8 $\alpha^+$  DCs have decreased mitochondrial fusion and/or excessive fission. Moreover, treatment of CD8 $\alpha^+$  DCs with mdivi-1 (a mitochondrial fission inhibitor) and M1 (a mitochondrial fusion promoter)<sup>14</sup> slightly enhanced the function of wild-type DCs, and notably, partially restored the ability of Mst1/2-deficient CD8 $\alpha^+$  DCs to prime CD8 $\alpha^+$  T cells (Fig. 3g). These results show that Mst1/2-dependent mitochondrial dynamics contribute to CD8 $\alpha^+$  DC function.

To identify additional mechanisms that regulate CD8 $\alpha^+$  DC function, we compared gene expression profiles of CD8 $\alpha^+$  or CD8 $\alpha^-$  DCs from wild-type and Mst1/2 $\Delta$ DC mice using gene-set enrichment analysis (GSEA). The majority of the pathways that were significantly upregulated owing to Mst1/2 deficiency were common to both CD8 $\alpha^+$  and CD8 $\alpha^-$  DCs (Fig. 4a and Extended Data Fig. 9a, b). By contrast, 18 pathways were significantly downregulated in Mst1/2-deficient CD8 $\alpha^+$  DCs, whereas none were significantly downregulated in CD8 $\alpha^-$  DCs (Fig. 4a and Extended Data Fig. 9c). IL-12 signalling was the most significantly downregulated gene set in Mst1/2-deficient CD8 $\alpha^+$  DCs (Fig. 4b and Extended Data Fig. 9c). *Il12b* (encoding IL-12 p40) expression was much higher in wild-type CD8 $\alpha^+$  DCs than in CD8 $\alpha^-$  DCs (Extended Data Fig. 9d), but this expression was considerably dampened by Mst1/2 deficiency (Fig. 4c). Additionally, IL-12 p40 or *Il12b* expression was diminished in CD8 $\alpha^+$  DCs upon deletion or inhibition of Mst1/2 (Extended Data Fig. 9e, f) and in FLT3L-cultured CD24<sup>high</sup> BMDCs lacking Mst1/2 (Extended Data Fig. 9g), whereas other cytokine genes were expressed normally (Extended Data Fig. 9h, i). Hence, Mst1/2 promote IL-12 expression in CD8 $\alpha^+$  DCs.

IL-12 serves as 'signal 3' for CD8 $\alpha^+$  T cell activation<sup>16</sup>, although its role in T cell homeostasis is less well understood. In *Il12a*<sup>-/-</sup> mice, there was a smaller effector/memory T cell population and fewer

IFN $\gamma$ -producing cells among CD8 $\alpha^+$  T cells but not among CD4 $\alpha^+$  T cells. This was also the case in Mst1/2 $\Delta$ DC and Mst1/2 $\Delta$ DC *Il12a*<sup>-/-</sup> mice (Extended Data Fig. 9j, k), highlighting critical roles of Mst1/2 and IL-12 in CD8 $\alpha^+$  T cell homeostasis. Moreover, *Il12a*<sup>-/-</sup> CD8 $\alpha^+$  DCs were impaired in priming CD8 $\alpha^+$  T cells (Extended Data Fig. 10a), as seen with Mst1/2-deficient cells. Notably, adding exogenous IL-12 to Mst1/2-deficient CD8 $\alpha^+$  DCs in vitro rescued, albeit incompletely, their ability to induce OT-I T cell proliferation (Fig. 4d), and IL-12 treatment in vivo promoted proliferation and CD44 upregulation in OT-I T cells in Mst1/2 $\Delta$ DC mice (Fig. 4e). Therefore, Mst1/2-mediated IL-12 expression contributes to CD8 $\alpha^+$  DC function in mediating CD8 $\alpha^+$  T cell homeostasis and priming.

How do Mst1 and Mst2 regulate IL-12 expression? Mst1/2-deficient CD8 $\alpha^+$  DCs exhibited reduced expression of NF- $\kappa$ B2 and RelB, which are involved in the non-canonical NF- $\kappa$ B pathway<sup>17</sup> (Fig. 4b, f and Extended Data Fig. 10b). A critical mediator of this pathway is NF- $\kappa$ B-inducing kinase (NIK), which functions in DCs to drive IL-12 production and CD8 $\alpha^+$  T-cell priming<sup>18</sup>. Ectopic expression of constitutively active NIK( $\Delta$ 78–84)<sup>19</sup> in FLT3L-cultured BMDCs promoted NF- $\kappa$ B2 and RelB expression in wild-type CD24<sup>high</sup> BMDCs (Extended Data Fig. 10c), and largely rescued the impaired expression of *Il12b* in Mst1/2-deficient cells (Fig. 4g). Moreover, expression of NIK( $\Delta$ 78–84) corrected the abnormal mitochondrial membrane potential of Mst1/2-deficient CD24<sup>high</sup> BMDCs (Extended Data Fig. 10d) and increased the ability of these cells to prime CD8 $\alpha^+$  T cells (Extended Data Fig. 10e). Further, Traf3, a key regulator of non-canonical NF- $\kappa$ B–NIK signalling<sup>17</sup>, interacted with Mst1 in CD8 $\alpha^+$  DCs (Extended Data Fig. 10f), although Traf3 expression was unaltered by Mst1/2 deficiency (Extended Data Fig. 10g). Collectively, Mst1 and Mst2 promote the non-canonical NF- $\kappa$ B pathway for IL-12 production and CD8 $\alpha^+$  DC function.

Given the roles of Mst1/2 in mitochondrial metabolism and IL-12 signalling, we examined whether mitochondrial metabolism is required for IL-12 production. Treatment of CD8 $\alpha^+$  DCs with metformin or oligomycin increased IL-12 expression (Extended Data Fig. 10h). Also, CD8 $\alpha^+$  DCs treated with exogenous IL-12 or deficient in IL-12 had

normal mitochondrial profiles (Extended Data Fig. 10i, j). However, NIK-deficient CD8 $\alpha^+$  DCs had abnormal mitochondrial membrane potential and mass (Extended Data Fig. 10k). Hence, mitochondrial profiles and IL-12 expression are largely discrete events that are coordinately regulated by Mst1/2, although defective non-canonical NF- $\kappa$ B signalling can alter mitochondrial homeostasis.

We explored regulation of Mst1/2 signalling in CD8 $\alpha^+$  DCs. FLT3L, which promotes DC development and expansion<sup>1</sup>, activated Mst1/2 and downstream signalling in CD8 $\alpha^+$  DCs (Extended Data Fig. 10l). Additionally, Mst1/2 deficiency impaired FLT3L-induced expansion of CD8 $\alpha^+$  DCs (Extended Data Fig. 10m), indicating a role of Mst1/2 in mediating FLT3L function. Next, we determined whether our findings in the mouse system were reflected in human cells. Mst1/2 inhibition altered mitochondrial mass and impaired IL-12 expression in human CD141 $^+$  DCs (Extended Data Fig. 10n, o), which are functionally equivalent to murine CD8 $\alpha^+$  DCs<sup>20</sup>. Therefore, Mst1/2 activity represents an evolutionarily conserved mechanism that regulates DC function.

By integrating a systems biology approach with experimental investigation, our work identifies Mst1/2 as crucial and selective regulators of CD8 $\alpha^+$  DC function and metabolism. Our NetBID algorithm has demonstrated its advantage over conventional methods by successfully identifying 'hidden' kinase drivers from high-throughput profiles, an approach with the potential to be extended to other drivers and biological questions. Mechanistically, we define a non-canonical Hippo signalling pathway (Mst1/2-dependent but Lats1/2- and Yap/Taz-independent) that coordinates mitochondrial activity and non-canonical NF- $\kappa$ B and cytokine signalling in CD8 $\alpha^+$  DCs (Extended Data Fig. 10p). These results highlight that the metabolic state and mitochondrial activity of DC subsets support their functional capacity, and point to a previously unappreciated mechanism that orchestrates DC subset function. Strategies that modulate the activities of DC subset-specific Mst1/2 signalling and metabolic regulation represent attractive means of therapeutic intervention against cancer and immune-mediated diseases.

## Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0177-0>.

Received: 4 July 2017; Accepted: 11 April 2018;

Published online: 30 May 2018

- Merad, M., Sathe, P., Helft, J., Miller, J. & Mortha, A. The dendritic cell lineage: ontogeny and function of dendritic cells and their subsets in the steady state and the inflamed setting. *Annu. Rev. Immunol.* **31**, 563–604 (2013).
- Murphy, K. M. Transcriptional control of dendritic cell development. *Annu. Rev. Immunol.* **34**, 93–119 (2016).
- Pearce, E. J. & Everts, B. Dendritic cell metabolism. *Nat. Rev. Immunol.* **15**, 18–29 (2015).
- O'Neill, L. A. J. & Pearce, E. J. Immunometabolism governs dendritic cell and macrophage function. *J. Exp. Med.* **213**, 15–23 (2016).
- Dudziak, D. et al. Differential antigen processing by dendritic cell subsets in vivo. *Science* **315**, 107–111 (2007).
- Piovan, E. et al. direct reversal of glucocorticoid resistance by AKT inhibition in acute lymphoblastic leukemia. *Cancer Cell* **24**, 766–776 (2013).

- Rodriguez-Barrueco, R. et al. Inhibition of the autocrine IL-6–JAK2–STAT3–calprotectin axis as targeted therapy for HR<sup>+</sup>/HER2<sup>+</sup> breast cancers. *Genes Dev.* **29**, 1631–1648 (2015).
- Huang, G. et al. Signaling via the kinase p38 $\alpha$  programs dendritic cells to drive T<sub>H</sub>17 differentiation and autoimmune inflammation. *Nat. Immunol.* **13**, 152–161 (2012).
- Meng, Z., Moroishi, T. & Guan, K.-L. Mechanisms of Hippo pathway regulation. *Genes Dev.* **30**, 1–17 (2016).
- Hildner, K. et al. Batf3 deficiency reveals a critical role for CD8 $\alpha^+$  dendritic cells in cytotoxic T cell immunity. *Science* **322**, 1097–1100 (2008).
- Mashayekhi, M. et al. CD8 $\alpha^+$  dendritic cells are the critical source of interleukin-12 that controls acute infection by *Toxoplasma gondii* tachyzoites. *Immunity* **35**, 249–259 (2011).
- Naik, S. H. et al. Cutting edge: generation of splenic CD8 $^+$  and CD8 $^-$  dendritic cell equivalents in fms-like tyrosine kinase 3 ligand bone marrow cultures. *J. Immunol.* **174**, 6592–6597 (2005).
- Cogliati, S. et al. Mitochondrial cristae shape determines respiratory chain supercomplexes assembly and respiratory efficiency. *Cell* **155**, 160–171 (2013).
- Buck, M. D. D. et al. Mitochondrial dynamics controls T cell fate through metabolic programming. *Cell* **166**, 63–76 (2016).
- Mishra, P. & Chan, D. C. Metabolic regulation of mitochondrial dynamics. *J. Cell Biol.* **212**, 379–387 (2016).
- Valenzuela, J., Schmidt, C. & Mescher, M. The roles of IL-12 in providing a third signal for clonal expansion of naive CD8 T cells. *J. Immunol.* **169**, 6842–6849 (2002).
- Sun, S. C. The non-canonical NF- $\kappa$ B pathway in immunity and inflammation. *Nat. Rev. Immunol.* **17**, 545–558 (2017).
- Katakam, A. K. et al. Dendritic cells require NIK for CD40-dependent cross-priming of CD8 $^+$  T cells. *Proc. Natl Acad. Sci. USA* **112**, 14664–14669 (2015).
- Sasaki, Y. et al. NIK overexpression amplifies, whereas ablation of its TRAF3-binding domain replaces BAF:BAFF-R-mediated survival signals in B cells. *Proc. Natl Acad. Sci. USA* **105**, 10883–10888 (2008).
- Bachem, A. et al. Superior antigen cross-presentation and XCR1 expression define human CD11c $^+$  CD141 $^+$  cells as homologues of mouse CD8 $^+$  dendritic cells. *J. Exp. Med.* **207**, 1273–1281 (2010).

**Acknowledgements** The authors acknowledge N. Chapman for critical reading of the manuscript, R. Johnson for *Stk4* and *Stk3* floxed mice, E. Olson for *Yap1* and *Taz* floxed mice, S.C. Sun for the NIK( $\Delta$ 78–84) construct, M. Hendren and A. KC for animal work, SJCRH Electron Microscopy core resource for electron microscopy, and Immunology FACS core facility for cell sorting. This work was supported by NIH AI105887, AI101407, CA176624, CA221290, NS064599 (to H.C.), and AG047928 (to J.P.).

**Reviewer information** Nature thanks L. O'Neill and the other anonymous reviewer(s) for their contribution to the peer review of this work.

**Author contributions** X.D. designed and performed in vitro and in vivo experiments, and wrote the manuscript; J.W. performed human DC study and immunoprecipitation; Y.W. initiated the project, and together with P.W.F.K. performed Seahorse assays; A.K. performed network inference; H.T., Y.L. and J.P. performed and analysed proteomics assays; C.G. performed STORM assays and helped with electron microscopy experiments; T.-L.M.N. performed the tumour study; Y.D. and G.N. performed functional enrichment analysis of microarrays; J.Y. developed the NetBID algorithm, performed systems biology analysis, wrote the manuscript and supervised overall computational analysis; and H.C. designed experiments, wrote the manuscript, and provided overall direction.

**Competing interests** The authors declare no competing interests.

## Additional information

**Extended data** is available for this paper at <https://doi.org/10.1038/s41586-018-0177-0>.

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41586-018-0177-0>.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>.

**Correspondence and requests for materials** should be addressed to J.Y. or H.C. **Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## METHODS

**Mice.** C57BL/6, CD45.1<sup>+</sup>, OT-I, OT-II, *Batf3*<sup>-/-</sup>, *B2m*<sup>-/-</sup>, *Il12a*<sup>-/-</sup>, *Mtor*<sup>fl</sup>, *Map3k14*<sup>-/-</sup>, *Lats1*<sup>fl</sup>, *Lats2*<sup>fl</sup>, and CD11c-Cre mice were purchased from The Jackson Laboratory. *Stk4*<sup>fl</sup> and *Stk3*<sup>fl</sup> mice were kindly provided by Randy Johnson<sup>21</sup>, and *Yap1*<sup>fl</sup> and *Taz*<sup>fl</sup> mice were kindly provided by E. Olson<sup>22</sup>. The mice have been backcrossed to the C57BL/6 background and were used at 8–12 weeks old. All of the genetically modified mice were viable and developed normally. For mixed bone marrow chimaera generation, bone marrow cells from wild-type or Mst1/2<sup>ΔDC</sup> CD45.2.2<sup>+</sup> mice were mixed with cells from CD45.1.2<sup>+</sup> mice at a 1:1 ratio and transferred into lethally irradiated (11 Gy) CD45.1.1<sup>+</sup> mice, followed by reconstitution for 6–8 weeks, as described previously<sup>23</sup>. In certain experiments, bone marrow cells from wild-type or Mst1/2<sup>ΔDC</sup> CD45.2.2<sup>+</sup> mice were transferred into lethally irradiated (11 Gy) CD45.1.1<sup>+</sup> mice. For chimaeras used in Fig. 2c, bone marrow cells from wild-type or Mst1/2<sup>ΔDC</sup> CD45.2.2<sup>+</sup> mice were mixed with bone marrow cells from *Batf3*<sup>-/-</sup> mice at a 1:1 ratio and transferred into lethally irradiated (11 Gy) CD45.1.1<sup>+</sup> mice. All mice were kept in a specific pathogen-free facility in the Animal Resource Center at St. Jude Children's Research Hospital. Animal protocols were approved by the Institutional Animal Care and Use Committee of St. Jude Children's Research Hospital.

**Cell purification.** Mouse spleens were digested with Collagenase D (Worthington) and CD11c<sup>+</sup> DCs were enriched using CD11c MicroBeads (Miltenyi Biotec) according to the manufacturer's instructions. Enriched cells were stained and sorted for CD8α<sup>+</sup> DCs (CD11c<sup>+</sup>CD8α<sup>+</sup>CD205<sup>+</sup>TCRβ<sup>+</sup>CD19<sup>+</sup>CD49b<sup>+</sup>B220<sup>-</sup>) and CD8α<sup>-</sup> DCs (CD11c<sup>+</sup>CD8α<sup>-</sup>CD205<sup>-</sup>TCRβ<sup>+</sup>CD19<sup>+</sup>CD49b<sup>+</sup>B220<sup>-</sup>) on a Reflection cell sorter (i-Cyt). Lymphocytes from spleen and peripheral lymph nodes were sorted for naive CD4<sup>+</sup> T cells (CD4<sup>+</sup>CD62L<sup>high</sup>CD44<sup>low</sup>CD25<sup>-</sup>) and naive CD8<sup>+</sup> T cells (CD8<sup>+</sup>CD62L<sup>high</sup>CD44<sup>low</sup>CD25<sup>-</sup>). The antibodies used for sorting were: anti-TCRβ-FITC (H57-597), anti-CD8-PE (53-6.7), anti-CD19-FITC (eBio1D3 (1D3)), anti-B220-FITC (RA3-6B2), anti-CD49b-FITC (DX5), anti-CD11c-PB (N418), anti-CD205-APC (205yekt), anti-CD4-PB (RM4-5), anti-CD25-FITC (PC61.5), anti-CD44-APC (IM7) and anti-CD62L-PE-Cy7 (MEL-14) (all from eBioscience); and anti-CD8-BV421 (53-6.7, Sony Biotechnology Inc.).

**Human DC isolation, real-time PCR and mitochondrial profile assays.** Peripheral blood mononuclear cells (PBMCs) were purified from blood using Ficoll (MP Biomedicals) and enriched for total DCs by immunodepletion. Briefly, Fc blocker (Miltenyi Biotec) and FITC-Lineage Cocktail 1 antibodies (BD Biosciences) were added to cells and incubated at 4°C for 20 min, followed by adding anti-mouse IgG MicroBeads and negative enrichment (Miltenyi Biotec). Flow-through was collected and stained with anti-HLA-DR-PE-Cy7 (LN3), anti-CD1c-PB (L161) (both from BioLegend), anti-CD141-BV605 (1A4, BD Biosciences), and anti-CD370-APC (8F9, Miltenyi Biotec), and Lin<sup>-</sup>HLA-DR<sup>+</sup>CD1c<sup>+</sup> cells (CD1c<sup>+</sup> DCs, equivalents to mouse CD8α<sup>-</sup> DCs) and Lin<sup>-</sup>HLA-DR<sup>+</sup>CD141<sup>+</sup>CD370<sup>+</sup> cells (CD141<sup>+</sup> DCs, equivalents to mouse CD8α<sup>+</sup> DCs) were sorted<sup>24</sup>. Sorted DCs were treated with vehicle or 1 μM Mst1/2 inhibitor (XMU-MP-1)<sup>25</sup> (Selleckchem) for 4 h, and RNA was extracted for reverse transcription and real-time PCR using SYBR Green (Thermo Fisher Scientific). The sequences for human IL-12 p40 primers were as follows: forward primer, GACATTCTGCGTTCAGGTCCAG; reverse primer, CATTTTTCGGCAGATGACCGTG. For mitochondrial mass analysis, DCs were treated with vehicle or 1.5 μM Mst1/2 inhibitor (XMU-MP-1)<sup>25</sup> for 4 h at 37°C and Mitotracker staining was performed according to the manufacturer's instructions (Invitrogen). All human studies were in compliance with the Declaration of Helsinki. Blood donors were recruited by the Blood Donor Center at St. Jude Children's Research Hospital. Blood donors provided written consent for their blood products not used in transfusions to be used for research. This consent form has been reviewed and approved by the Institutional Review Board at St. Jude Children's Research Hospital.

**In vitro bone marrow-derived DCs.** Bone marrow cells were flushed from mouse tibias and femurs, and red blood cells were lysed using ACK lysis buffer. For FLT3L-BMDCs, bone marrow cells were cultured in RPMI-1640 medium (plus β-mercaptoethanol) containing 10% (vol/vol) FBS and 1% (vol/vol) penicillin-streptomycin (complete RPMI-1640 medium) and mouse FLT3L (200 ng/ml) for 9.5–11.5 days. For GM-CSF-BMDCs, bone marrow cells were cultured in complete RPMI-1640 medium and mouse GM-CSF (20 ng/ml) and IL-4 (5 ng/ml) for 7.5–9.5 days. Loosely adherent cells were collected for analysis.

**Antigen challenge.** Antigen-specific T cells from OT-I (CD45.1<sup>+</sup>; 1 × 10<sup>6</sup>) TCR-transgenic mice were sorted and labelled with CFSE, and transferred into wild-type and Mst1/2<sup>ΔDC</sup> mice intravenously. Twenty-four hours later, the mice were injected intravenously with 30 μg OVA (low Endo, Worthington). Three days after OVA immunization, spleen and peripheral lymph nodes were isolated for further analyses. For in vivo IL-12 treatment, 2 μg IL-12 per mouse was injected intraperitoneally each day for three consecutive days starting from 24 h after OT-I CD8<sup>+</sup> T cell transfer.

**Tumour model.** MC38 colon adenocarcinoma cells were cultured in DMEM supplemented with 10% (vol/vol) FBS and 1% (vol/vol) penicillin-streptomycin. Wild-type and Mst1/2<sup>ΔDC</sup>-derived bone marrow chimaera mice were injected subcutaneously with 5 × 10<sup>5</sup> MC38 cells in the right flank. Tumours were measured regularly with calipers. Tumour volumes were calculated using the formula: length × width × width × π/6. To prepare tumour-infiltrating lymphocytes, tumour tissues were excised, minced and digested with 0.5 mg/ml Collagenase IV (Roche) + 200 U/ml DNase I (Sigma) for 1 h at 37°C. Tumour-infiltrating lymphocytes were then isolated by density-gradient centrifugation over Percoll (Life Technologies). **Listeria monocytogenes-OVA infection model.** Mice were infected intravenously with 3 × 10<sup>4</sup> colony-forming units (CFU) of *L. monocytogenes* expressing chicken ovalbumin (LM-OVA), and were bled at the indicated days post infection to collect cells for analysis.

**Flow cytometry.** Flow cytometry was performed as described previously<sup>23</sup>, with the following antibodies: anti-IL-4-APC (11B11), anti-H-2K<sup>b</sup>-APC (AF6-88.5.5.3), anti-CD70-PE (FR70), anti-TCRβ-APC-Cy7 (H57-597), anti-IFNγ-PE-Cy7 (XMG1.2), anti-CD44-APC (IM7), anti-CD62L-PE-Cy7 (MEL-14), anti-CD44-APC-Cy7 (IM7), anti-CD44-FITC (IM7), anti-CD11c-PE (N418), anti-CD205-APC (205yekt), anti-CD40-PE (1C10), anti-MHC-II-APC-Cy7 (M5/114.15.2), anti-CD80-APC (16-10A1), anti-CD86-FITC (GL1), anti-4-1BBL-PE (TKS-1), anti-OX40L-APC (RM134L), anti-PD-L1-PE (MIH5), anti-LAG3-PE (eBioC9B7W) and anti-CD11c-PE-Cy7 (N418) (all from eBioscience); anti-CD44-BV650 (IM7), anti-CD11b-BV650 (M1/70), anti-IL-17A-BV421 (TC11), anti-PD-1-BV421 (29F.1A12), anti-TIM3-APC (B8.2c12) and anti-CD4-BV711 (RM4-5) (all from Biolegend); anti-IL-2-PE (JES6-5H4, BD Biosciences); and anti-CD8-BV605 (53-6.7, Sony Biotechnology Inc.). For intracellular cytokine detection, cells were stimulated for 4 to 5 h with phorbol 12-myristate 13-acetate (PMA) and ionomycin in the presence of monensin before staining according to the manufacturer's instructions (BD Biosciences). Mitotracker and tetramethyl-rhodamine, methyl ester (TMRM) staining was performed according to the manufacturer's instructions (Invitrogen). Flow cytometry data were acquired on LSRII or LSR Fortessa (BD Biosciences) and analysed using FlowJo software (Tree Star).

**Antigen presentation assays.** For in vitro assays, CD8α<sup>+</sup> and CD8α<sup>-</sup> DCs were sorted from spleen, pulsed with 500 μg/ml OVA, 250 pg/ml OVA (257–264) or 3 μg/ml OVA (323–339) peptide for 1 h, then washed twice and cultured with naive CD4<sup>+</sup> T cells from OT-II mice or naive CD8<sup>+</sup> T cells from OT-I mice for three days. Thymidine was added to the culture 7–8 h before cells were collected to measure proliferation. Where indicated, DCs were pre-treated with 0.2 μM M1<sup>26</sup> + 0.1 μM mdivi-1<sup>27</sup> for 4 h before co-culture with T cells. For in vivo assays by DC transfer method, FLT3L (10 μg/ml in 100 μl PBS) was injected subcutaneously once daily into wild-type or Mst1/2<sup>ΔDC</sup> mice for nine consecutive days<sup>28</sup>, and then CD8α<sup>+</sup> and CD8α<sup>-</sup> DCs were sorted and pulsed with 20 mg/ml OVA for 1.5 h, washed and intravenously injected into *B2m*<sup>-/-</sup> mice that had received CFSE-labelled CD8<sup>+</sup> OT-I (CD45.1<sup>+</sup>) T cells the day before. *B2m*<sup>-/-</sup> mice were killed three days after DC immunization, and CFSE dilution of CD8<sup>+</sup> OT-I T cells was examined by flow cytometry. In vivo cross-presentation assay using tetramer detection methods was performed according to previous reports<sup>29,30</sup>. In brief, *B2m*<sup>-/-</sup> splenocytes were osmotically loaded with 10 mg/ml OVA (Worthington Biochemical Corporation), washed, irradiated at 13.5 Gy, and 1.5 × 10<sup>7</sup> cells were intravenously injected into mice. After eight days, spleens were harvested and analysed by H-2K<sup>b</sup>-OVA tetramer staining.

**RNA and immunoblot analyses.** Real-time PCR analysis was performed as previously described with primers and probe sets from Applied Biosystems<sup>31</sup>. Immunoblots were performed and quantified as described previously<sup>12</sup>, using the following antibodies: NF-κB2 (#4882), RelB (#4954), p-Mst1/2 (#3681), p-Yap (Ser127) (D9W21), Lats1 (C66B5), Mst1 (D8B9Q), Mst2 (#3952), Actin (8H10D10), p-DRP1 (Ser637) (#4867), p-S6 (2F9), c-Myc (#9402 s) (all from Cell Signaling Technology), OPA1 (NB110-55290SS, Novus Biologicals) and MitoProfile Total OXPHOS (oxidative phosphorylation) Rodent WB Antibody Cocktail (MS604, MitoSciences).

**Co-immunoprecipitation of Mst1-associated complexes from splenic CD8α<sup>+</sup> DCs.** Wild-type splenic CD8α<sup>+</sup> DCs (after in vivo FLT3L expansion to obtain sufficient cells) were lysed in lysis buffer (25 mM Tris-HCl pH 7.5, 130 mM NaCl, 20 mM NaF, 1% Triton X-100, 2 mM EDTA) and pre-cleaned by incubating with 40 μl Protein A/G beads (50% vol/vol slurry, Santa Cruz sc-2003) for 2 h. The Mst1-associated complexes were immunoprecipitated with Mst1 antibody (EP1465Y, Abcam), which had pre-coupled with Protein A/G Sepharose beads for 4 h at 4°C. After three washes with the lysis buffer, the immune complexes were analysed by immunoblotting with PKA Cα/β (#515741, R&D), Traf3 (#4729, Cell Signaling Technology), and Mst1 antibodies.

**Super-resolution fluorescence microscopy.** Stochastic optical reconstruction microscopy (STORM) was performed as described<sup>32</sup>. In brief, sorted DC subsets were seeded onto Poly-L-Lysine coated chamber slides (Ibidi USA) and allowed to settle for 30 min before fixation with 4% paraformaldehyde followed by treatment

with sodium borohydride to quench free reactive groups. Cells were permeabilized for 3 min in buffer (50 mM Tris-HCl, pH 8.0, 50 mM NaCl) containing 0.1% Triton X-100 before blocking in buffer containing 2% BSA and 0.05% Tween-20. Samples were incubated overnight in TBS buffer containing BSA and anti-Tom20 (Santa Cruz; sc-11415, 1:500 dilution) and subsequently detected with CF647 labelled secondary antibody (Biotium). Three-dimensional STORM image acquisition was performed with an N-STORM system (Nikon Instruments) comprised of a  $100\times$  1.49NA TIRF objective and an astigmatic lens inserted into the light path, before collection using an iXon DU897 ultra EMCCD camera with frame rate of 109 frames per second and EM gain of 17 MHz at 16 bit, as previously described<sup>32</sup>. Images were processed using algorithms as previously described<sup>33</sup> and incorporated into Elements imaging software (Nikon Instruments).

**Transmission electron microscopy.** Sorted DC subsets were pelleted by centrifugation and fixed in 0.1 M sodium cacodylate buffer, pH 7.4, containing 2.5% glutaraldehyde and 2% paraformaldehyde. Samples were post-fixed in 2% osmium tetroxide in 0.1 M cacodylate buffer with 0.15% potassium ferrocyanide, followed by dehydration to propylene oxide and embedding in epoxy resin. TEM images of ultrathin (80 nm) sections were acquired using a FEI Tecnai 20 200 KV FEG electron microscope.

**Metabolic assays.** Oxygen consumption rates (OCR) and extracellular acidification rates (ECAR) were measured in Seahorse XF media under basal conditions and in response to 1  $\mu$ M oligomycin, 1.5  $\mu$ M fluoro-carbonyl cyanide phenylhydrazide (FCCP) and 500 nM rotenone using an XF96 Extracellular Flux Analyzer (EFA) (Seahorse Bioscience).

**Whole and phosphoproteome profiling by multiplex TMT-LC/LC-MS/MS.** *Protein extraction, digestion, labelling and pooling.* Whole proteome and phosphoproteome profiling was performed as described<sup>34</sup>. CD8 $\alpha^+$  and CD8 $\alpha^-$  DCs were sorted from wild-type spleen as described above. Cells were washed twice with ice-cold PBS and cell pellets from six samples ( $n=3$  per cell type) were lysed in fresh lysis buffer (50 mM HEPES, pH 8.5, 8 M urea and 0.5% sodium deoxycholate). The protein concentration of lysate was quantified by BCA protein assay (Thermo Fisher Scientific). Proteins (100  $\mu$ g) from each sample were loaded and run on a 10% SDS-PAGE gel until all samples were inside the gel. The gel was stained and each sample was sliced and further chopped into 1 mm<sup>3</sup> pieces. After gel detaining, the proteins were reduced with 5 mM dithiothreitol at 37 °C for 30 min and alkylated with 10 mM iodoacetamide at room temperature in the dark for 30 min. Proteins were then digested in-gel with trypsin in 50 mM HEPES overnight at 37 °C with a protein to trypsin ratio (w/w) of 50:1. Peptides for each sample were extracted, dried and labelled with 6-plex tandem mass tag (TMT, Thermo Fisher Scientific) reagents following the manufacturer's instructions. Finally, the TMT-labelled samples were equally mixed.

*Offline basic pH reverse phase liquid chromatography.* The mixture of the 6 TMT-labelled samples was desalted, dried and solubilized in 60  $\mu$ l buffer A (10 mM ammonium formate, pH 8) and separated on an XBridge C18 column (3.5- $\mu$ m particle size, 4.6 mm  $\times$  25 cm, Waters) into 44 fractions with an 88-min gradient from 15% to 45% buffer B (95% acetonitrile, 10 mM ammonium formate, pH 8, flow rate of 0.4 ml/min). Each fraction was dried for whole-proteome analysis.

*Acidic pH reverse phase liquid chromatography coupled with tandem mass spectrometry.* The analysis was performed based on our optimized platform<sup>35</sup>. For whole proteome analysis, the dried peptides were reconstituted in 5% formic acid, loaded on a reverse phase column (75  $\mu$ m  $\times$  30 cm, 1.9  $\mu$ m C18 resin (Dr. Maisch, Germany)) interfaced with a Q-Exactive HF mass spectrometer (ThermoFisher Scientific). Peptides were eluted by 12–36% buffer B gradient in 2.5 h (buffer A: 0.2% formic acid, 3% DMSO; buffer B: buffer A plus 67% acetonitrile, flow rate of 0.25  $\mu$ l/min). The column was heated at 65 °C by a butterfly portfolio heater (Phoenix S&T) to reduce back pressure. The mass spectrometer was operated in data-dependent mode with a survey scan in Orbitrap (60,000 resolution,  $1\times 10^6$  AGC target and 50 ms maximal ion time) and 20 tandem mass spectrometry (MS/MS) high-resolution scans (60,000 resolution,  $1\times 10^5$  AGC target, 105 ms maximal ion time, HCD, 35 normalized collision energy, 1.0  $m/z$  isolation window, and 20 s dynamic exclusion).

**Proteomics data analysis.** The analysis was performed by our in-house JUMP search engine which has been used for data processing in previous publications<sup>34</sup>. In brief, acquired MS/MS raw files were converted into mzXML format and searched by the JUMP algorithm against a composite target/decoy database to estimate FDR. The target protein database was downloaded from the Uniprot mouse database (52,490 protein entries) and the decoy protein database was generated by reversing all target protein sequences. Searches were performed with 10 ppm mass tolerance for both precursor ions and product ions, fully tryptic restriction, two maximal missed cleavages and the assignment of  $a$ ,  $b$  and  $y$  ions. TMT tags on lysine residues and peptide N termini (+229.162932 Da) and carbamidomethylation of cysteine residues (+57.021 Da) were used for static modifications and oxidation of methionine residues (+15.99492 Da) were used for dynamic modification. The assigned peptides were filtered by mass accuracy, minimal peptide

length, matching scores, charge state and trypticity to reduce protein false discovery rate (FDR) to below 1%.

**TMT-based protein quantification.** The analysis was performed using in-house JUMP software suite as previously reported<sup>34</sup>. In brief, TMT reporter ion intensities of each peptide spectrum match (PSM) were extracted and PSMs with very low intensity were removed. The raw intensities were then corrected based on isotopic distribution of each labelling reagent and loading bias. The mean-centred intensities across samples were calculated and protein relative intensities were derived by averaging related PSMs. Finally, protein absolute intensities were determined by multiplying the relative intensities by the grand-mean of three most highly abundant PSMs.

**Gene expression profiling and bioinformatics analysis.** Total, CD8 $\alpha^+$  and CD8 $\alpha^-$  DCs ( $n=4$  per genotype) were isolated from spleen as described above. RNA was obtained with an RNeasy Micro Kit according to the manufacturer's instructions (Qiagen). RNA samples were then analysed with the Mouse Gene 2.0 ST Signals array. Differentially expressed transcripts were identified by ANOVA (Partek Genomics Suite version 6.5), and the Benjamini–Hochberg method was used to estimate the FDR as described<sup>36</sup>. GSEA was performed as described<sup>36</sup>.

**Batch effect removal of combined gene expression data.** We combined multiple batches of microarray gene expression profiles of total ( $n=15$ ), CD8 $\alpha^+$  ( $n=4$ ) and CD8 $\alpha^-$  ( $n=4$ ) DCs, and removed batch effects by using 'removeBatchEffect' function in limma<sup>37</sup>. A principal component analysis (PCA) plot of the corrected microarray profiles (Extended Data Fig. 1a) indicated significant differences among the three groups of DCs.

**NetBID algorithm.** We and others have demonstrated that important signalling proteins might not change at individual mRNA expression levels<sup>6,7,38</sup>, and existing network-based methods<sup>6,7,38–41</sup> to infer master regulators are largely based on transcriptomics data only. In order to identify the true underlying signalling drivers of CD8 $\alpha^+$  DCs in an unbiased and systematic manner, we developed the data-driven network-based Bayesian inference of drivers (NetBID) algorithm by integrating transcriptomics (mRNA), whole proteomics (wProtein) and phosphoproteomics (pProtein) data. First, we collected a cohort of baseline transcriptomic profiles of total DCs and reverse engineered a DC-specific signalling interactome or DCI, using an improved version of ARACNE<sup>42</sup>, an information theory-based algorithm for regulatory network inference. The data-driven DCI resulted in 20,846 nodes and 660,929 edges. Second, we focused on kinase signalling networks in DCI and calculated activity scores of all kinase candidates ( $n=289$  present in all three platforms) in mRNA, wProtein and pProtein profiles of CD8 $\alpha^+$  and CD8 $\alpha^-$  DCs using  $z$ -normalization and  $z$ -statistic<sup>43,44</sup>. We then used a Bayesian linear modelling approach<sup>45</sup> to identify differentially activated kinases by comparing profiles of CD8 $\alpha^+$  with CD8 $\alpha^-$  DCs at mRNA, wProtein and pProtein levels separately. Finally, we integrated three differential activity scores using the unsigned Stouffer's method<sup>46</sup> and identified 36 kinase drivers that were differentially activated between CD8 $\alpha^+$  and CD8 $\alpha^-$  DCs at mRNA, wProtein and pProtein levels with cutoffs of FDR at 0.01 and network size at 50. Remarkably, most of the 36 kinases themselves showed little change between CD8 $\alpha^+$  and CD8 $\alpha^-$  DCs or even reduced in CD8 $\alpha^+$  DCs, and would likely be missed by conventional differential expression approaches. Mst1 and a few other Hippo kinases were confirmed to have higher activity in CD8 $\alpha^+$  DCs, indicating the power of NetBID to identify 'hidden' drivers.

**Signalling pathway enrichment of identified kinase drivers.** To identify any known signalling pathways enriched by the 36 NetBID-inferred kinase drivers of CD8 $\alpha^+$  DCs, we used Fisher's exact test against all pathways with 'signaling' in their names from MSigDB<sup>47</sup> (v6.0). We manually curated the Hippo signalling kinases using the most recent literature<sup>9</sup> as we observed a number of known Hippo kinases present in the 36 kinase driver list, which turned out to be more significant than all existing signalling pathways in the database. This also suggests the limitation of knowledge-based pathway databases.

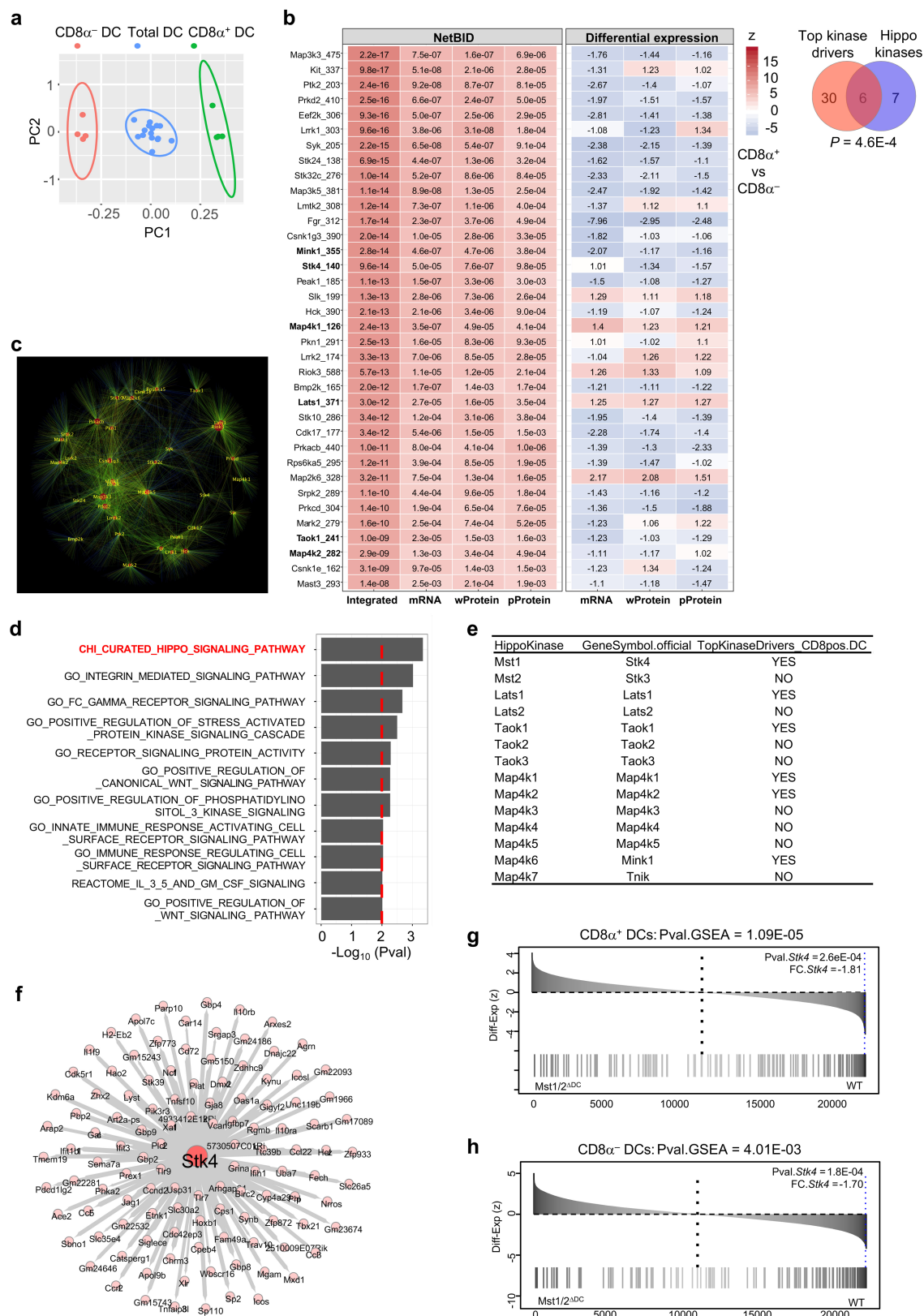
**Retroviral generation and transduction of FLT3L-BMDCs.** Retroviruses were produced by transfection of Plat-E cells with an empty pCLXSN (GFP) vector or the same vector encoding NIK( $\Delta$ 78–84) mutant, along with pCL-Eco packaging vectors. NIK( $\Delta$ 78–84) lacks a Traf3-binding motif and is potent in activating the non-canonical NF- $\kappa$ B pathway<sup>19</sup>. Bone marrow cells were cultured with FLT3L for one day; cells were then 'spin-infected' (2,500 r.p.m.) for 180 min at 30 °C. After spinning, fresh DC medium containing FLT3L was added to the cells and cells were cultured according to the standard procedure described above.

**Statistical analysis for small-scale immunological experiments.** The experiments were not randomized, and investigators were not blinded to allocation during experiments and outcome assessment. Data were analysed using Prism 6 software (GraphPad) by two-tailed Student's  $t$ -test or one-way or two-way ANOVA as noted in figure legends.  $P<0.05$  was considered significant. Data are presented as mean  $\pm$  s.e.m.

**Reporting summary.** Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

**Software and data availability.** The NetBID software package is available at <https://github.com/jyyulab/NetBID>. Microarray data are available via Gene Expression Omnibus under accession number GSE100772. Proteomics data are available via ProteomeXchange (<http://www.proteomexchange.org/>) with identifier PXD006875.

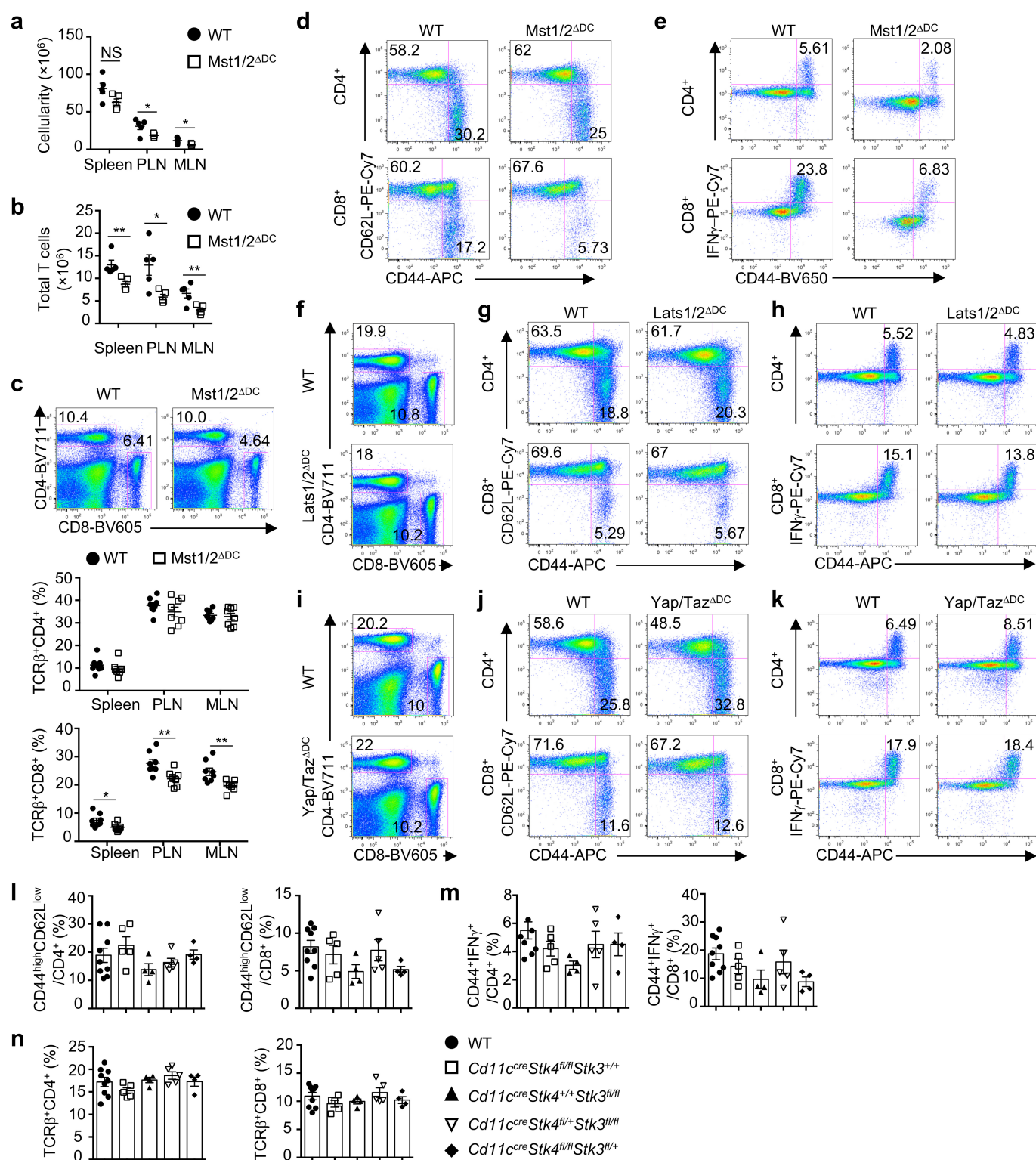
21. Lu, L. et al. Hippo signaling is a potent in vivo growth and tumor suppressor pathway in the mammalian liver. *Proc. Natl Acad. Sci. USA* **107**, 1437–1442 (2010).
22. Xin, M. et al. Hippo pathway effector Yap promotes cardiac regeneration. *Proc. Natl Acad. Sci. USA* **110**, 13839–13844 (2013).
23. Wang, Y. et al. Transforming growth factor beta-activated kinase 1 (TAK1)-dependent checkpoint in the survival of dendritic cells promotes immune homeostasis and function. *Proc. Natl Acad. Sci. USA* **109**, E343–E352 (2012).
24. Jongbloed, S. L. et al. Human CD141<sup>+</sup> (BDCA-3)<sup>+</sup> dendritic cells (DCs) represent a unique myeloid DC subset that cross-presents necrotic cell antigens. *J. Exp. Med.* **207**, 1247–1260 (2010).
25. Fan, F. et al. Pharmacological targeting of kinases MST1 and MST2 augments tissue repair and regeneration. *Sci. Transl. Med.* **8**, 352ra108 (2016).
26. Wang, D. et al. A small molecule promotes mitochondrial fusion in mammalian cells. *Angew. Chem. Int. Ed.* **51**, 9302–9305 (2012).
27. Cassidy-Stone, A. et al. Chemical inhibition of the mitochondrial division dynamin reveals its role in Bax/Bak-dependent mitochondrial outer membrane permeabilization. *Dev. Cell* **14**, 193–204 (2008).
28. Maraskovsky, E. et al. Dramatic increase in the numbers of functionally mature dendritic cells in Flt3 ligand-treated mice: multiple dendritic cell subpopulations identified. *J. Exp. Med.* **184**, 1953–1962 (1996).
29. den Haan, J. M., Lehar, S. M. & Bevan, M. J. CD8<sup>+</sup> but not CD8<sup>−</sup> dendritic cells cross-prime cytotoxic T cells in vivo. *J. Exp. Med.* **192**, 1685–1696 (2000).
30. Kretzer, N. M. et al. RAB43 facilitates cross-presentation of cell-associated antigens by CD8 $\alpha$ <sup>+</sup> dendritic cells. *J. Exp. Med.* **213**, 2871–2883 (2016).
31. Wang, Y. et al. Tuberous sclerosis 1 (Tsc1)-dependent metabolic checkpoint controls development of dendritic cells. *Proc. Natl Acad. Sci. USA* **110**, E4894–E4903 (2013).
32. Liedmann, S. et al. Viral suppressors of the RIG-I-mediated interferon response are pre-packaged in influenza virions. *Nat. Commun.* **5**, 5645 (2014).
33. Huang, B., Wang, W., Bates, M. & Zhuang, X. Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy. *Science* **319**, 810–813 (2008).
34. Tan, H. et al. Integrative proteomics and phosphoproteomics profiling reveals dynamic signaling networks and bioenergetics pathways underlying T cell activation. *Immunity* **46**, 488–503 (2017).
35. Wang, H. et al. Systematic optimization of long gradient chromatography mass spectrometry for deep analysis of brain proteome. *J. Proteome Res.* **14**, 829–838 (2015).
36. Zeng, H. et al. mTORC1 couples immune signals and metabolic programming to establish T<sub>reg</sub>-cell function. *Nature* **499**, 485–490 (2013).
37. Ritchie, M. E. et al. *limma* powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
38. Putcha, P. et al. HDAC6 activity is a non-oncogene addiction hub for inflammatory breast cancers. *Breast Cancer Res.* **17**, 149 (2015).
39. Wang, K. et al. Genome-wide identification of post-translational modulators of transcription factor activity in human B cells. *Nat. Biotechnol.* **27**, 829–837 (2009).
40. Alvarez, M. J. et al. Functional characterization of somatic mutations in cancer using network-based inference of protein activity. *Nat. Genet.* **48**, 838–847 (2016).
41. Lefebvre, C. et al. A human B-cell interactome identifies MYB and FOXM1 as master regulators of proliferation in germinal centers. *Mol. Syst. Biol.* **6**, 377 (2010).
42. Margolin, A. A. et al. ARACNE: An algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* **7**, S7 (2006).
43. Efron, B. & Tibshirani, R. On testing the significance of sets of genes. *Ann. Appl. Stat.* **1**, 107–129 (2007).
44. Irizarry, R. A., Wang, C., Zhou, Y. & Speed, T. P. Gene set enrichment analysis made simple. *Stat. Methods Med. Res.* **18**, 565–575 (2009).
45. Gelman, A. et al. *Bayesian Data Analysis* 3rd edn. (CRC Press, Boca Raton, 2013).
46. Riley, J. W. et al. The American soldier: adjustment during army life. *Am. Sociol. Rev.* **14**, 557–559 (1949).
47. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).



Extended Data Fig. 1 | See next page for caption.

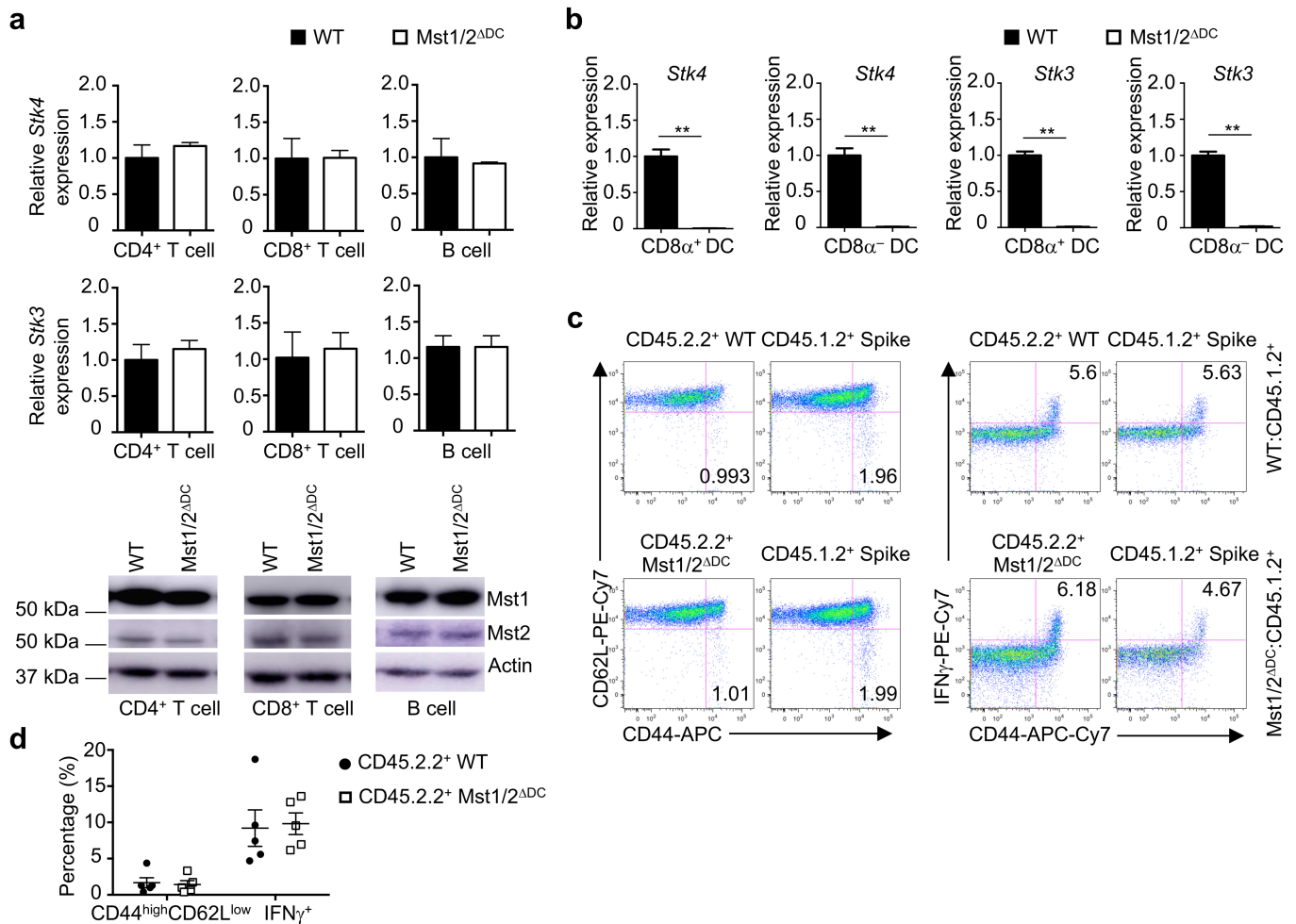
**Extended Data Fig. 1 | NetBID analysis for the reconstruction of the DC signalling interactome (DCI), network and enrichment analyses of top kinase drivers, and identification and validation of *Stk4* (*Mst1*) regulons.** **a**, PCA plot of baseline microarray gene expression profiles of total DCs (blue,  $n = 15$ ; used for de novo DCI reconstruction), CD8 $\alpha^+$  and CD8 $\alpha^-$  DCs (green and red, respectively,  $n = 4$  each; used for differential expression analysis) after removal of batch effects. **b**, Top 36 hub kinases that are differentially activated in CD8 $\alpha^+$  DCs relative to CD8 $\alpha^-$  DCs inferred by NetBID. Left, the NetBID panel indicates the significance level (colour coded by  $z$  score; labelled values are  $P$  values) of the driver network in integrated analysis, transcriptomics (mRNA), whole proteomics (wProtein) and phosphoproteomics (pProtein) data, respectively. Right, differential expression of the drivers (colour coded by  $z$  score; labelled values are signed fold changes). The Venn diagram shows the enrichment

of Hippo pathway kinases in the top putative kinase drivers. **c**, Network interactions of top 36 kinase drivers of CD8 $\alpha^+$  DCs. **d**, Top signalling pathways enriched by 36 NetBID-inferred kinase drivers ( $P < 0.01$ , number of overlapped genes  $> 2$ ). **e**, Known kinases in Hippo signalling<sup>9</sup> and analysis by NetBID. **f**, *Stk4*-mediated gene network ( $n = 140$ ) from DCI computationally inferred from baseline gene expression profiles of total DCs by NetBID. The width of an edge is proportional to the pairwise mutual information of connected nodes. **g**, **h**, Enrichment of predicted *Mst1* signalling regulons (as shown in **f**) in differentially expressed genes between *Mst1/2*-deficient (*Mst1/2* $^{\Delta DC}$ ) and wild-type CD8 $\alpha^+$  DCs (**g**) or CD8 $\alpha^-$  DCs (**h**). Pval.GSEA indicates the  $P$  value of GSEA; Pval.*Stk4* and FC.*Stk4* indicate the  $P$  value and signed fold change of *Stk4* expression (insert).



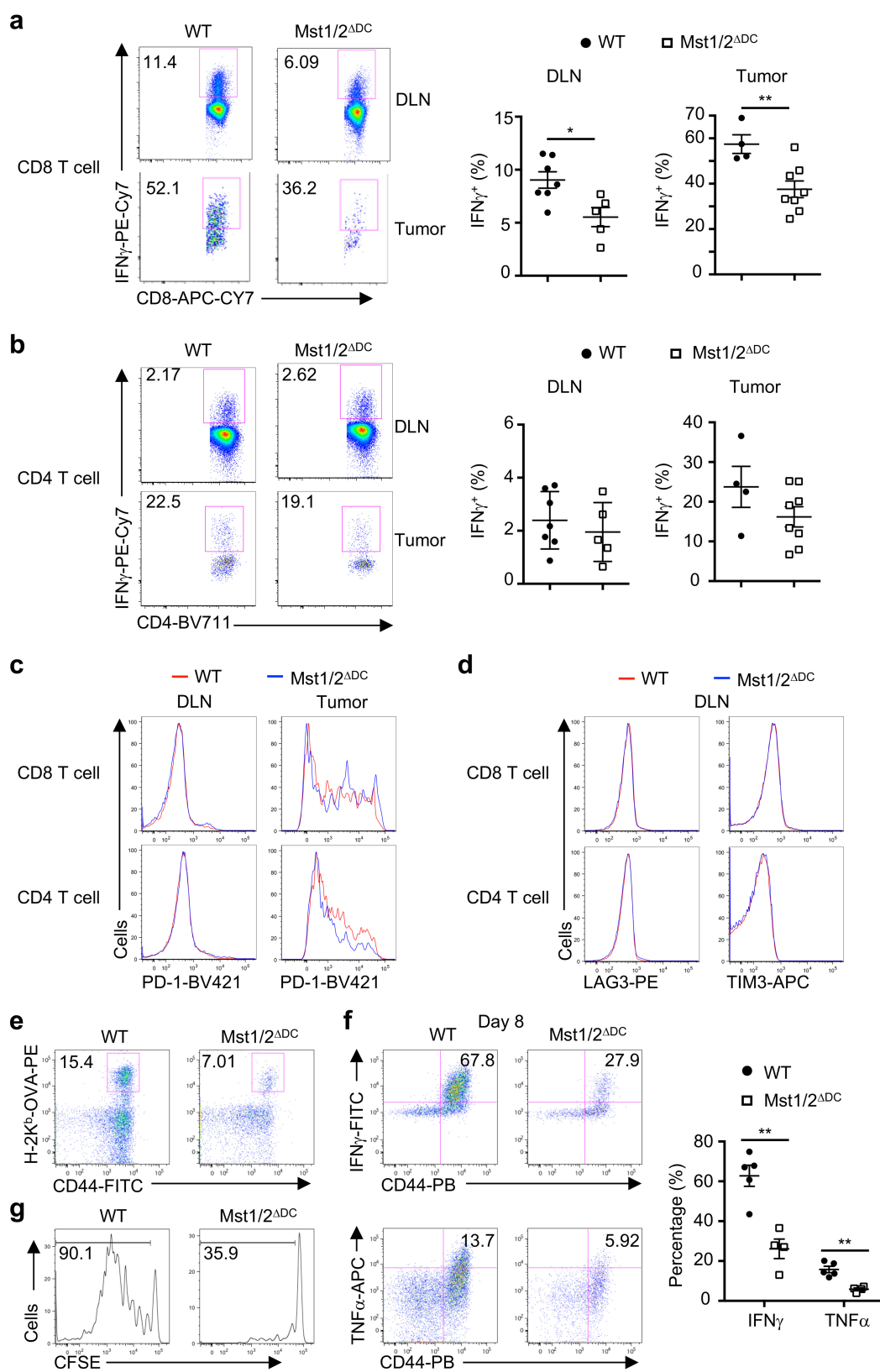
**Extended Data Fig. 2 | T cell homeostasis in mice with DC-specific deletion of Hippo pathway genes.** **a, b**, Total cellularity (**a**) or T cell numbers (TCR $\beta^+$  CD8 $^+$  and TCR $\beta^+$  CD4 $^+$ ) (**b**) of the spleen, peripheral lymph nodes (PLN) and mesenteric lymph nodes (MLN) of wild-type and Mst1/2 $\Delta$ DC mice ( $n = 5$  mice per genotype). **c**, Flow cytometry analysis of splenic CD4 $^+$  and CD8 $^+$  T cell populations (upper) and frequencies of total CD4 $^+$  and CD8 $^+$  T cells in spleen, PLN and MLN (lower) of wild-type and Mst1/2 $\Delta$ DC mice ( $n = 8$  mice per genotype). **d**, CD44 and CD62L expression on splenic CD4 $^+$  and CD8 $^+$  T cells of wild-type and Mst1/2 $\Delta$ DC mice. **e**, CD44 and IFN $\gamma$  expression in splenic CD4 $^+$  and CD8 $^+$  T cells. **f–k**, Flow cytometry analysis of CD4 $^+$  and CD8 $^+$  populations, expression of CD44 and CD62L, or CD44

and IFN $\gamma$  in CD4 $^+$  and CD8 $^+$  T cells from spleen of wild-type and Lats1/2 $\Delta$ DC (**f–h**) or Yap/Taz $\Delta$ DC mice (**i–k**). **l–n**, Frequencies of CD44 $^{\text{high}}$ CD62L $^{\text{low}}$  effector/memory cells (**l**) and CD44 $^+$ IFN $\gamma$  $^+$  cells (**m**) in splenic CD4 $^+$  and CD8 $^+$  T cells and frequencies of splenic CD4 $^+$  and CD8 $^+$  T cells (**n**) of wild-type ( $n = 9$ ), Cd11c $^{\text{cre}}$ Stk4 $^{\text{fl/fl}}$ Stk3 $^{+/+}$  ( $n = 5$ ), Cd11c $^{\text{cre}}$ Stk4 $^{+/+}$ Stk3 $^{\text{fl/fl}}$  ( $n = 4$ ), Cd11c $^{\text{cre}}$ Stk4 $^{\text{fl/fl}}$ Stk3 $^{\text{fl/fl}}$  ( $n = 5$ ) and Cd11c $^{\text{cre}}$ Stk4 $^{\text{fl/fl}}$ Stk3 $^{\text{fl/fl}}$  ( $n = 4$ ) mice (Cd11c is also known as *Itgax*). Numbers in quadrants or gates indicate percentage of cells. Data are mean and s.e.m. NS, not significant; \* $P < 0.05$ ; \*\* $P < 0.01$ ; two-tailed unpaired Student's *t*-test in a–c; one-way ANOVA in l–n. Data summarize two (i–k), three (f–h), four (a, b, d, e), six (c) or eight (l–n) independent experiments.



**Extended Data Fig. 3 | Analyses of Mst1 and Mst2 deletion in DCs and lymphocytes from Mst1/2<sup>ΔDC</sup> mice and T cell homeostatic status in mixed bone marrow chimaeras.** **a**, Real-time PCR (upper and middle) and immunoblot (lower) analyses of *Stk4* (also known as *Mst1*) and *Stk3* (also known as *Mst2*) mRNA and protein expression in CD4<sup>+</sup> T cells, CD8<sup>+</sup> T cells and B cells from wild-type and Mst1/2<sup>ΔDC</sup> mice ( $n = 3$  mice per genotype). **b**, Real-time PCR analysis of *Stk4* and *Stk3* mRNA expression in splenic CD8<sup>α</sup><sup>+</sup> and CD8<sup>α</sup><sup>-</sup> DCs from wild-type and Mst1/2<sup>ΔDC</sup> mice ( $n = 3$  for accessing Mst2 expression in Mst1/2-deficient CD8<sup>α</sup><sup>+</sup> DCs,  $n = 4$  for others). **c**, **d**, Bone marrow cells from wild-type or

Mst1/2<sup>ΔDC</sup> CD45.2.2<sup>+</sup> mice were mixed with cells from CD45.1.2<sup>+</sup> (spike) mice at a 1:1 ratio and transferred into lethally irradiated CD45.1.1<sup>+</sup> mice. After 6–8 weeks, bone marrow chimaeras were analysed for the expression of CD44, CD62L and IFN<sub>γ</sub> (**c**) and frequencies of CD44<sup>high</sup> CD62L<sup>low</sup> effector/memory cells and CD44<sup>+</sup> IFN<sub>γ</sub><sup>+</sup> cells (**d**) in splenic CD8<sup>+</sup> T cells derived from wild-type or Mst1/2<sup>ΔDC</sup> donor bone marrow cells (CD45.2.2<sup>+</sup>) or spike cells CD45.1.2<sup>+</sup> ( $n = 5$  mice per genotype). Numbers in quadrants indicate percentage of cells. Data are shown as mean and s.e.m. \*\* $P < 0.01$ ; two-tailed unpaired Student's *t*-test in **a**, **b**, **d**. Data summarize two (**a**, **b**) or three (**c**, **d**) independent experiments.

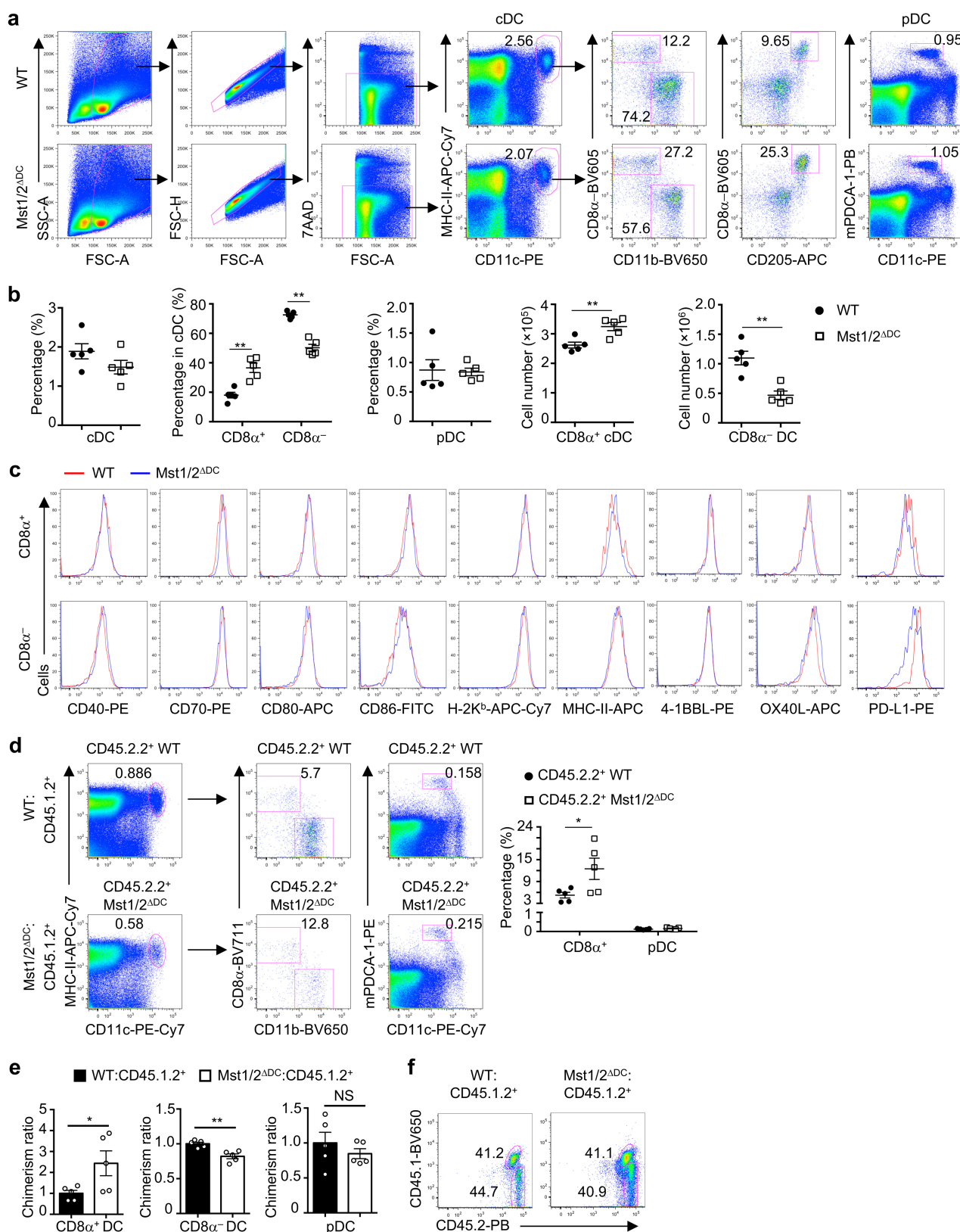


Extended Data Fig. 4 | See next page for caption.

**Extended Data Fig. 4 | In vivo T cell responses in wild-type and Mst1/2<sup>ΔDC</sup> mice challenged with tumour, pathogen or cognate antigen.**

**a, b**, Flow cytometry analysis of IFN $\gamma$  expression (left) and frequencies of IFN $\gamma$ <sup>+</sup> cells (right) in CD8<sup>+</sup> (**a**) and CD4<sup>+</sup> (**b**) T cells from draining lymph node (DLN) and tumour tissues of wild-type and Mst1/2<sup>ΔDC</sup> mice challenged with MC38 tumour cells ( $n = 7$  for wild type,  $n = 5$  for Mst1/2<sup>ΔDC</sup> for DLN;  $n = 4$  for wild type,  $n = 8$  for Mst1/2<sup>ΔDC</sup> for tumour tissues). **c**, Flow cytometry analysis of PD-1 expression on CD8<sup>+</sup> (upper) and CD4<sup>+</sup> (lower) T cells from DLN and tumour tissues of wild-type and Mst1/2<sup>ΔDC</sup> mice challenged with MC38 tumour cells. **d**, Flow cytometry analysis of LAG3 and TIM3 expression of CD8<sup>+</sup> (upper) and CD4<sup>+</sup>

(lower) T cells from DLN of wild-type and Mst1/2<sup>ΔDC</sup> mice challenged with MC38 tumour cells. **e**, Flow cytometry of H-2K<sup>b</sup>-OVA<sup>+</sup> CD8<sup>+</sup> T cells in blood from wild-type and Mst1/2<sup>ΔDC</sup> mice infected with LM-OVA. **f**, Flow cytometry (left) and frequencies (right) of IFN $\gamma$ <sup>+</sup> and TNF $\alpha$ <sup>+</sup> cells of PMA and ionomycin-stimulated CD8<sup>+</sup> T cells in the blood from wild-type and Mst1/2<sup>ΔDC</sup> mice infected with LM-OVA ( $n = 5$  for wild type,  $n = 4$  for Mst1/2<sup>ΔDC</sup>). **g**, CFSE dilution of donor OT-I T cells in OVA-immunized mice. Numbers in quadrants or gates indicate percentage of cells. Data are shown as mean and s.e.m. \* $P < 0.05$ , \*\* $P < 0.01$ ; two-tailed unpaired Student's  $t$ -test in **a**, **b**, **f**. Data summarize two (**a–e**, **g**) independent experiments.

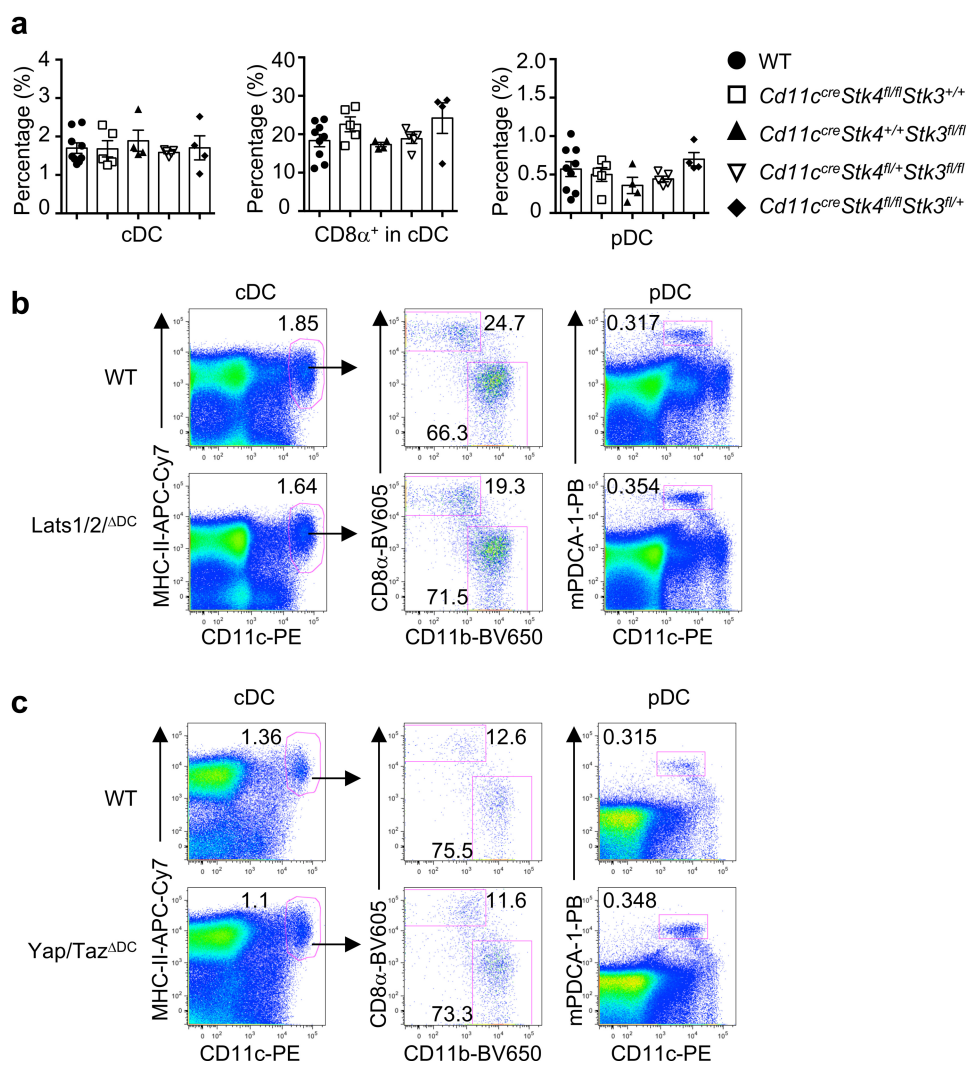


Extended Data Fig. 5 | See next page for caption.

**Extended Data Fig. 5 | Altered homeostasis of Mst1/2-deficient DCs.**

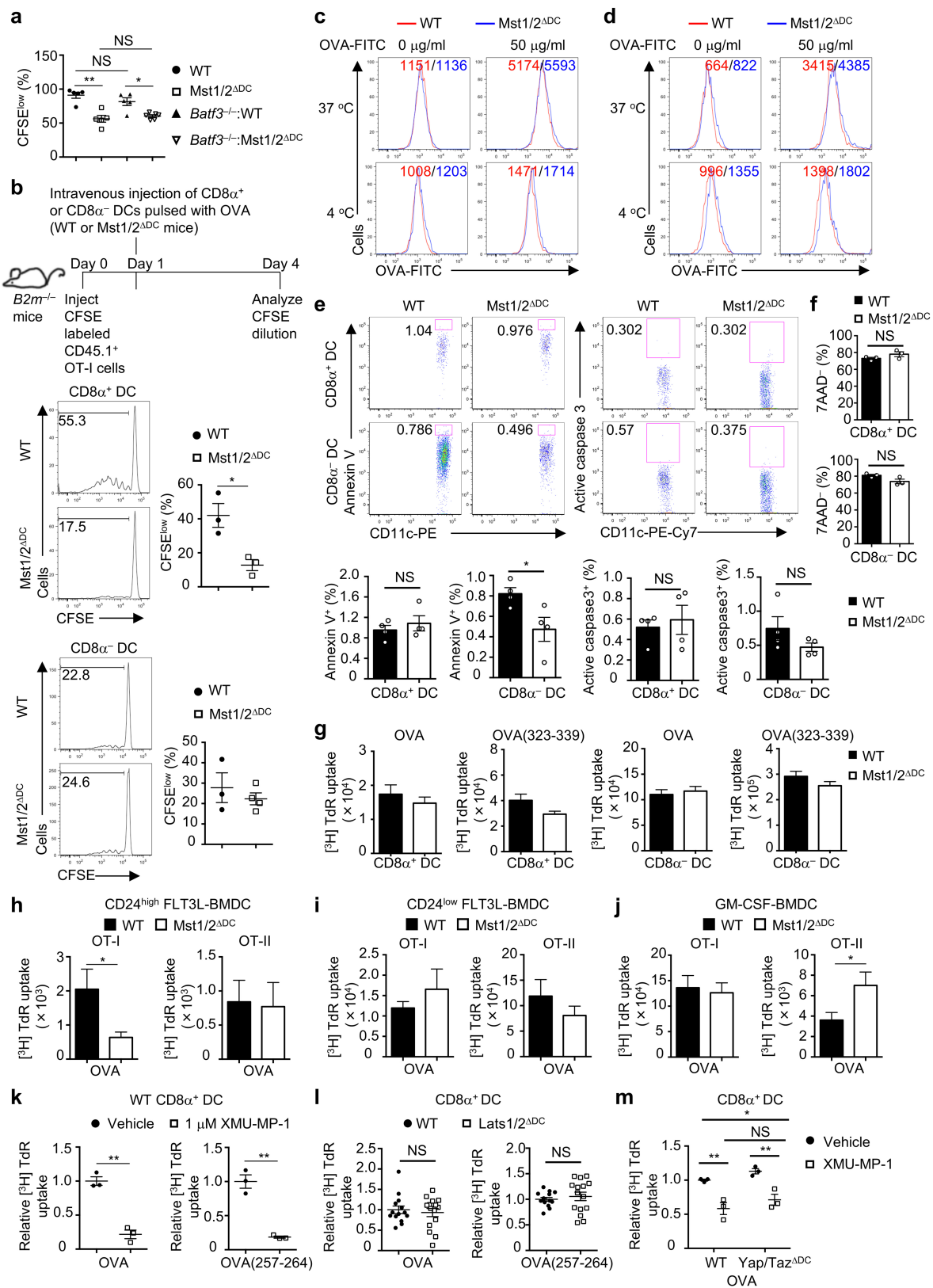
**a**, Detailed gating strategy for flow cytometry of splenic conventional DCs (cDC), CD8 $\alpha^+$  cDC (CD8 $\alpha^+$  CD11b $^-$ ), CD8 $\alpha^-$  cDC (CD8 $\alpha^-$  CD11b $^+$ ) and pDC populations in wild-type and Mst1/2 $^{\Delta DC}$  mice. MHC-II, MHC class II. **b**, Frequencies and cell numbers of splenic DC populations in wild-type and Mst1/2 $^{\Delta DC}$  mice as gated in **a** ( $n = 5$  mice per genotype). **c**, Flow cytometry analysis of CD40, CD70, CD80, CD86, H-2K $^b$ , MHC-II, 4-1BBL, OX40L and PD-L1 expression on splenic CD8 $\alpha^+$  and CD8 $\alpha^-$  DCs as gated in **a**. **d**, Flow cytometry (left) of CD45.1.2 $^+$  bone marrow cell-derived and CD45.2.2 $^+$  wild-type or Mst1/2 $^{\Delta DC}$  donor bone marrow cell-derived splenic cDC, CD8 $\alpha^+$  cDC, CD8 $\alpha^-$  cDC and pDC populations in mixed chimaeras and frequencies (right) of CD8 $\alpha^+$  cDCs in total cDCs

and pDCs in spleen derived from wild-type or Mst1/2 $^{\Delta DC}$  donor bone marrow cells in mixed chimaeras ( $n = 5$  mice per genotype). **e**, Normalized chimaerism for the indicated DC subsets in mixed bone marrow chimaeras. The percentage of indicated DC subsets was normalized by that of B cells from same mice. The chimaerism of the wild type was set as 1 ( $n = 5$  mice per genotype). **f**, Flow cytometry analysis of donor (wild-type or Mst1/2 $^{\Delta DC}$ , CD45.2.2 $^+$ ) and spike (CD45.1.2 $^+$ ) bone marrow cell percentages in the bone marrow mixture before transfer. Numbers in gates indicate percentage of cells. Data are shown as mean and s.e.m. \* $P < 0.05$ , \*\* $P < 0.01$ ; two-tailed unpaired Student's  $t$ -test in **b**, **d**, **e**. Data summarize four (**a–c**) or three (**d**, **e**) independent experiments.



**Extended Data Fig. 6 | Homeostasis of DCs after deletion of Hippo pathway genes.** **a**, Frequencies of splenic cDC, CD8 $\alpha^+$  cDC, and pDC populations in wild-type ( $n = 9$ ), *Cd11c<sup>cre</sup>Stk4<sup>fl/fl</sup>Stk3<sup>+/+</sup>* ( $n = 5$ ), *Cd11c<sup>cre</sup>Stk4<sup>+/+</sup>Stk3<sup>fl/fl</sup>* ( $n = 4$ ), *Cd11c<sup>cre</sup>Stk4<sup>fl/+</sup>Stk3<sup>fl/fl</sup>* ( $n = 5$ ) and *Cd11c<sup>cre</sup>Stk4<sup>fl/fl</sup>Stk3<sup>fl/+</sup>* ( $n = 4$ ) mice. **b**, **c**, Flow cytometry of splenic

cDC, CD8 $\alpha^+$  cDC, CD8 $\alpha^-$  cDC and pDC populations in wild-type and *Lats1/2 $\Delta$ DC* (**b**), or *Yap/Taz $\Delta$ DC* (**c**) mice. Numbers in gates indicate percentage of cells. Data summarize two (**b**), three (**c**) or eight (**a**) independent experiments.

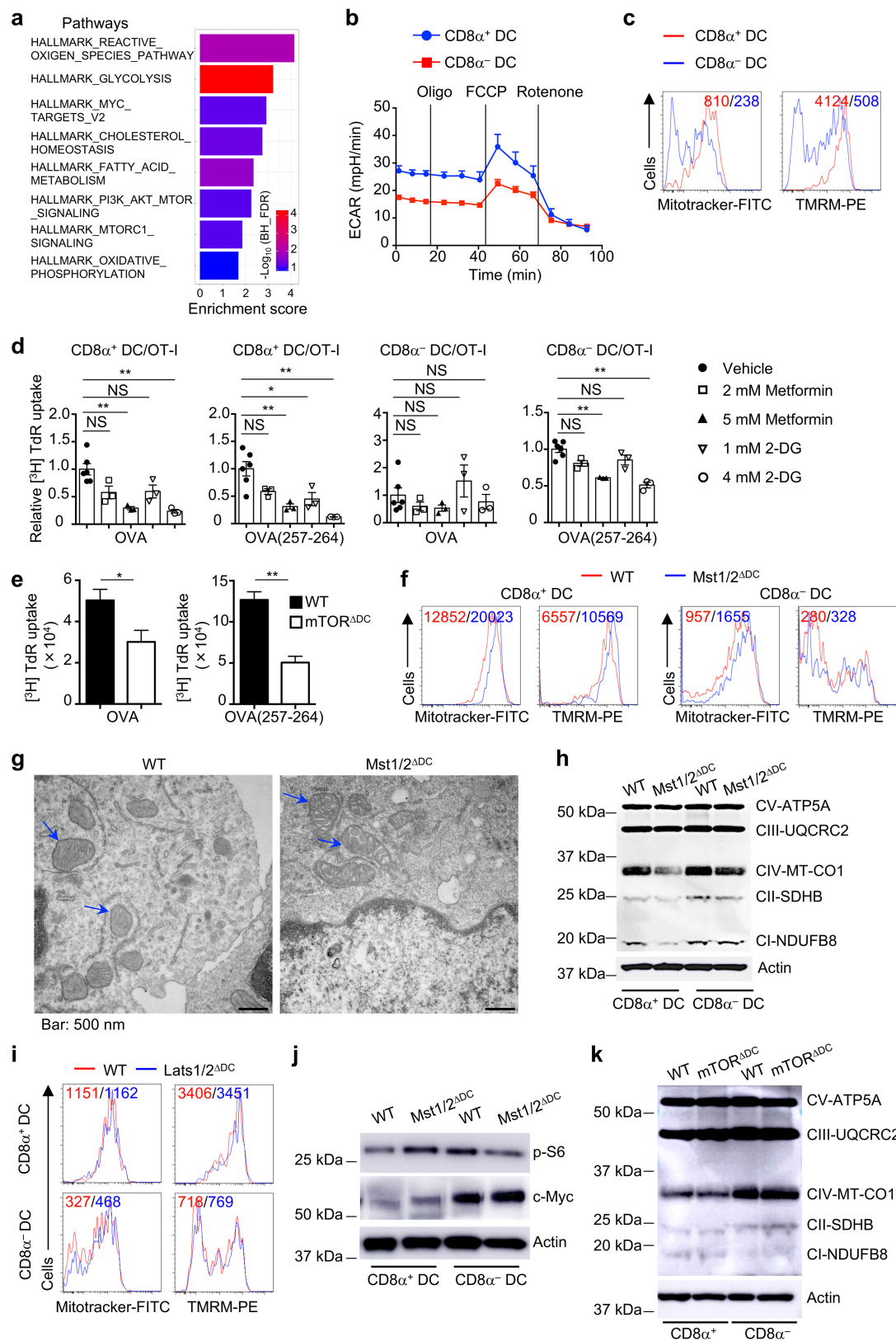


Extended Data Fig. 7 | See next page for caption.

**Extended Data Fig. 7 | Role of Mst1/2 in selectively programming functions of CD8 $\alpha^+$  DCs and CD24<sup>high</sup> FLT3L-BMDCs to prime CD8 T cells.**

**a**, Frequency of CFSE<sup>low</sup> cells of donor OT-I T cells in wild-type ( $n = 5$ ), Mst1/2<sup>ΔDC</sup> ( $n = 5$ ), *Batf3*<sup>-/-</sup>:wild-type ( $n = 5$ ) or *Batf3*<sup>-/-</sup>:Mst1/2<sup>ΔDC</sup> ( $n = 6$ ) mixed chimaeras immunized with OVA. **b**, CFSE-labelled OT-I T cells (CD45.1<sup>+</sup>) were transferred to *B2m*<sup>-/-</sup> mice followed by immunization one day later with OVA-pulsed CD8 $\alpha^+$  or CD8 $\alpha^-$  splenic DCs isolated from wild-type or Mst1/2<sup>ΔDC</sup> mice (following DC expansion by FLT3L), and CFSE dilution of OT-I cells was examined three days after DC immunization. Shown are representative flow cytometry histograms of CFSE dilution and frequency of CFSE<sup>low</sup> cells in donor OT-I T cells (gated on CD45.1<sup>+</sup>CD45.2<sup>-</sup>) from spleen of *B2m*<sup>-/-</sup> mice ( $n = 4$  for Mst1/2-deficient CD8 $\alpha^-$  DCs,  $n = 3$  for all others). **c, d**, CD8 $\alpha^+$  (**c**) or CD8 $\alpha^-$  DCs (**d**) from wild-type and Mst1/2<sup>ΔDC</sup> mice were fed with 0 or 50 μg/ml soluble OVA conjugated with FITC for 1 h at 37 °C or 4 °C. Cells were then collected and OVA uptake was evaluated by flow cytometry. Numbers in graphs indicate the mean fluorescence intensity of OVA-FITC. **e**, Flow cytometry (upper) and quantification (lower,  $n = 4$  per genotype) of apoptotic CD8 $\alpha^+$  or CD8 $\alpha^-$  DCs examined by Annexin V (left) and active caspase 3 (right) staining in freshly isolated splenocytes from wild-type and Mst1/2<sup>ΔDC</sup> mice. **f**, CD8 $\alpha^+$  (upper) or CD8 $\alpha^-$  DCs (lower) from wild-type and Mst1/2<sup>ΔDC</sup> mice were cultured overnight for analysis of cell viability by 7-aminoactinomycin D (7-AAD) staining. Quantification of the percentage of 7-AAD-negative live cells is shown ( $n = 3$  per genotype). **g**, Thymidine incorporation of OT-II T cells cultured with OVA protein- or OVA(323–339) peptide-pulsed CD8 $\alpha^+$

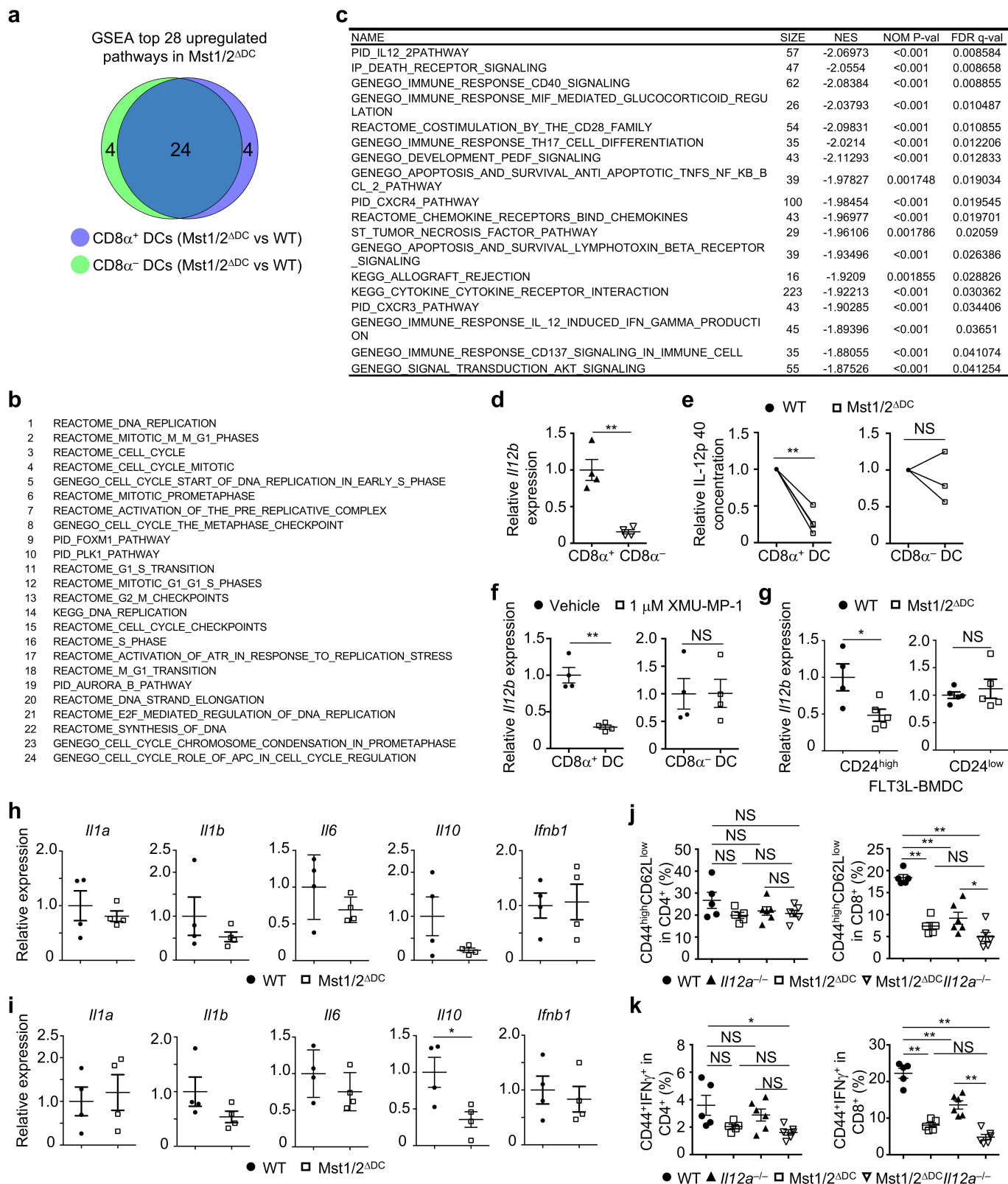
or CD8 $\alpha^-$  DCs ( $n = 13$  derived from five mice for the Mst1/2-deficient CD8 $\alpha^+$  DC/OVA(323–339) group,  $n = 15$  derived from five mice for all other groups). **h–j**, Thymidine incorporation of OT-I (left) or OT-II (right) T cells cultured with OVA protein-pulsed CD24<sup>high</sup> FLT3L-BMDCs (**h**), CD24<sup>low</sup> FLT3L-BMDCs (**i**), or GM-CSF-derived BMDCs (**j**) from wild-type and Mst1/2<sup>ΔDC</sup> mice for 72 h ( $n = 14$  from five mice for the Mst1/2-deficient GM-CSF-BMDC/OT-I group,  $n = 15$  from five mice for all other groups). **k**, Relative thymidine incorporation of OT-I T cells cultured with OVA protein- or OVA(257–264) peptide-pulsed splenic CD8 $\alpha^+$  DCs pre-treated with vehicle or Mst1/2 inhibitor (XMU-MP-1) for 4 h. Thymidine incorporation of OT-I T cells cultured with vehicle-treated DCs in each group was set as 1 ( $n = 3$  per group). **l**, Relative thymidine incorporation of OT-I T cells cultured with OVA protein- or OVA(257–264) peptide-pulsed CD8 $\alpha^+$  DCs from WT and *Lats1/2*<sup>ΔDC</sup> mice ( $n = 14$  derived from five mice for wild type,  $n = 15$  derived from five mice for *Lats1/2*<sup>ΔDC</sup>). Thymidine incorporation of OT-I T cells cultured with wild-type DCs in each group was set as 1. **m**, Relative thymidine incorporation of OT-I T cells cultured with OVA protein-pulsed wild-type or *Yap/Taz*<sup>ΔDC</sup> splenic CD8 $\alpha^+$  DCs that were pre-treated with vehicle or Mst1/2 inhibitor ( $n = 3$  derived from two mice per genotype). Thymidine incorporation of OT-I T cells cultured with vehicle-treated DCs was set as 1. Numbers in gates indicate percentage of cells. Data are shown as mean and s.e.m. \* $P < 0.05$ , \*\* $P < 0.01$ ; two-tailed unpaired Student's *t*-test in **a, b, e–l**; two-way ANOVA in **m**. Data summarize two (**a–d, g–j**), three (**e, f, k, l**) or four (**e**) independent experiments.



Extended Data Fig. 8 | See next page for caption.

**Extended Data Fig. 8 | Analysis of mitochondrial profiles of Mst1/2-, Lats1/2- or mTOR-deficient DCs.** **a**, Functional annotations of upregulated metabolic pathways according to KEGG and Hallmark databases in CD8 $\alpha^+$  DCs (compared to CD8 $\alpha^-$  DCs) profiled using proteomics. **b**, ECAR of splenic CD8 $\alpha^+$  and CD8 $\alpha^-$  DCs. Oligo, oligomycin; FCCP, carbonyl cyanide *p*-trifluoromethoxyphenylhydrazide. **c**, Flow cytometry analysis of mitochondrial mass and membrane potential of wild-type splenic CD8 $\alpha^+$  and CD8 $\alpha^-$  DCs using Mitotracker and TMRM (Tetramethylrhodamine, methyl ester) staining, respectively. Numbers in graph indicate the mean fluorescence intensity. **d**, Relative thymidine incorporation of OT-I T cells cultured with OVA protein- or OVA(257–264) peptide-pulsed CD8 $\alpha^+$  and CD8 $\alpha^-$  DCs pre-treated with vehicle or metabolic inhibitors. Values after culture with vehicle-treated DCs were set as 1. **e**, Thymidine incorporation of OT-I T cells cultured with OVA protein- or OVA(257–264) peptide-pulsed splenic CD8 $\alpha^+$  DCs from wild-type and mTOR $^{\Delta DC}$  mice for 72 h ( $n = 12$  from four mice per genotype). **f**, Flow cytometry analysis of mitochondrial mass and mitochondrial membrane potential of wild-type and Mst1/2 $^{\Delta DC}$

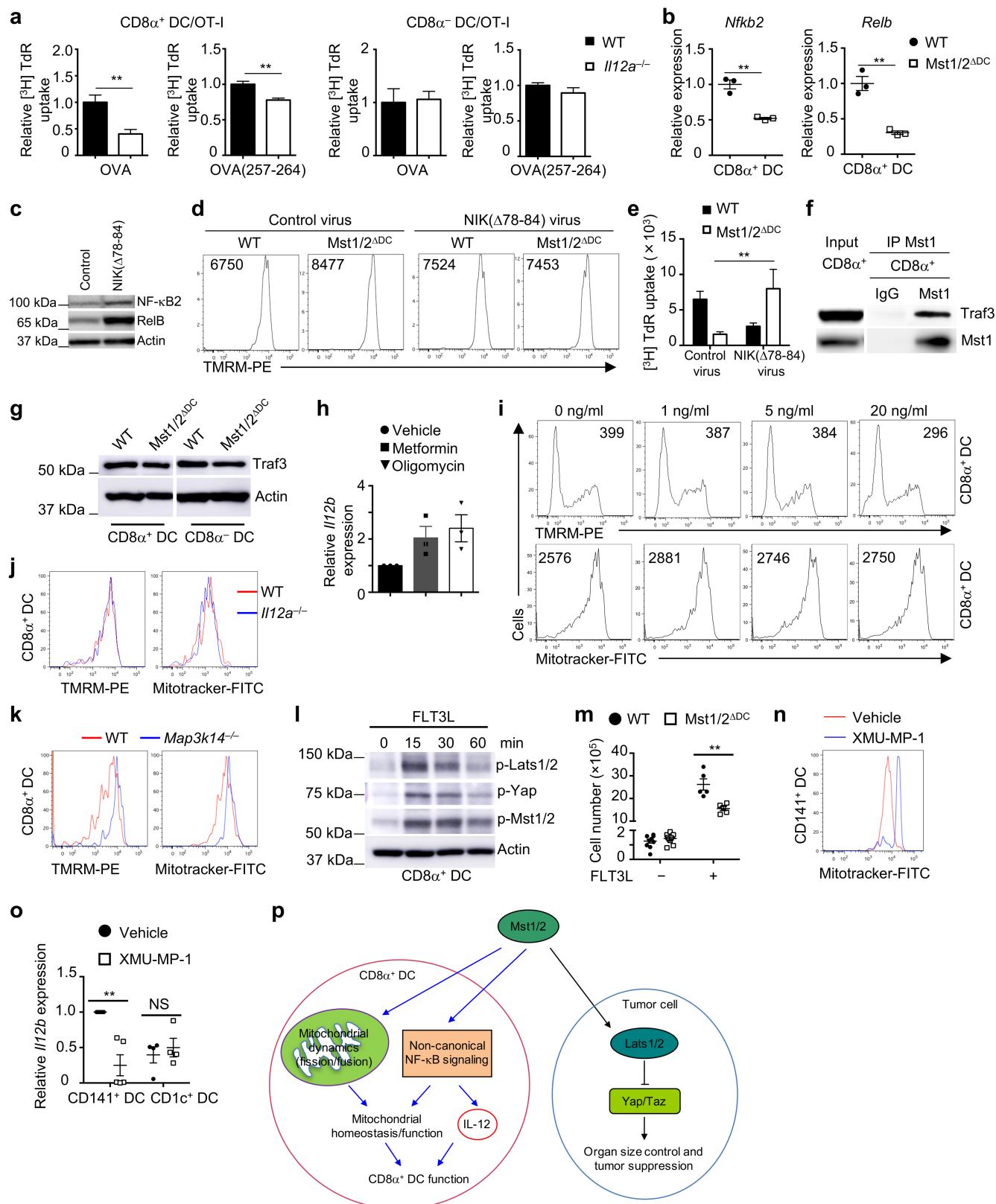
splenic CD8 $\alpha^+$  and CD8 $\alpha^-$  DCs by Mitotracker and TMRM staining, respectively. Numbers in graph indicate the mean fluorescence intensity. **g**, Transmission electron microscopy analysis of mitochondria of splenic CD8 $\alpha^-$  DCs from wild-type or Mst1/2 $^{\Delta DC}$  mice. Arrows indicate mitochondria. **h**, Immunoblot analysis of expression of NDUFB8 (complex I), SDHB (complex II), UQCRC2 (complex III), MT-CO1 (complex IV) and ATP5A (complex V) in CD8 $\alpha^+$  and CD8 $\alpha^-$  DCs. **i**, Flow cytometry analysis of mitochondrial mass and mitochondrial membrane potential of wild-type and Lats1/2 $^{\Delta DC}$  splenic CD8 $\alpha^+$  and CD8 $\alpha^-$  DCs by Mitotracker and TMRM staining, respectively. Numbers in graph indicate the mean fluorescence intensity. **j**, Immunoblot analysis of p-S6 and c-Myc protein in CD8 $\alpha^+$  and CD8 $\alpha^-$  DCs of wild-type and Mst1/2 $^{\Delta DC}$  mice. **k**, Immunoblot analysis of NDUFB8 (complex I), SDHB (complex II), UQCRC2 (complex III), MT-CO1 (complex IV) and ATP5A (complex V) protein in CD8 $\alpha^+$  and CD8 $\alpha^-$  DCs of wild-type and mTOR $^{\Delta DC}$  mice. Data are shown as mean and s.e.m. \* $P < 0.05$ , \*\* $P < 0.01$ ; one-way ANOVA in **d**; two-tailed unpaired Student's *t*-test in **e**. Data summarize two (**e**, **f**, **h–k**), three (**b**, **d**) or four (**c**) independent experiments.



Extended Data Fig. 9 | See next page for caption.

**Extended Data Fig. 9 | Selective regulation of IL-12 signalling and expression by Mst1/2 in CD8 $\alpha^+$  DCs.** **a**, Venn diagram showing the overlap of top 28 upregulated (Mst1/2 $\Delta^{DC}$  versus wild type) pathways by GSEA between CD8 $\alpha^+$  and CD8 $\alpha^-$  DCs. Briefly, transcriptional profiles of Mst1/2-deficient CD8 $\alpha^+$  and CD8 $\alpha^-$  DCs were compared to their respective wild-type counterparts by GSEA, and then upregulated pathways (FDR < 0.05) were identified in CD8 $\alpha^+$  DCs (Mst1/2 $\Delta^{DC}$  versus wild type) and CD8 $\alpha^-$  DCs (Mst1/2 $\Delta^{DC}$  versus wild type) and used to generate a Venn diagram. The top 28 upregulated pathways in Mst1/2-deficient CD8 $\alpha^+$  and CD8 $\alpha^-$  DCs (compared to their respective wild-type counterparts) were largely shared (24/28) between the two DC subsets. **b**, List of the significantly upregulated (FDR < 0.05) 24 pathways (out of 28) shared by Mst1/2-deficient CD8 $\alpha^+$  and CD8 $\alpha^-$  DCs (compared to their respective wild-type counterparts), as revealed by GSEA. **c**, List of the significantly downregulated (FDR < 0.05) pathways determined by GSEA (arranged according to FDR values) in Mst1/2-deficient CD8 $\alpha^+$  DCs (versus wild-type cells). NES, normalized enrichment score. **d**, *Il12b* expression in wild-type CD8 $\alpha^+$  and CD8 $\alpha^-$  DCs. **e**, Relative IL-12 p40

cytokine concentration in the supernatant of lipopolysaccharide (LPS)-treated CD8 $\alpha^+$  and CD8 $\alpha^-$  DCs from wild-type and Mst1/2 $\Delta^{DC}$  mice ( $n = 4$  per genotype for CD8 $\alpha^+$  DCs,  $n = 3$  per genotype for CD8 $\alpha^-$  DCs). IL-12 p40 cytokine concentration of wild-type DCs was set as 1. **f**, Real-time PCR analysis of *Il12b* mRNA expression in wild-type splenic CD8 $\alpha^+$  and CD8 $\alpha^-$  DCs treated with vehicle or Mst1/2 inhibitor (XMU-MP-1) ( $n = 4$  per treatment). **g**, *Il12b* expression in FLT3L-BMDCs ( $n = 4$  for wild-type CD24 $^{high}$  BMDCs,  $n = 5$  for other groups). **h**, **i**, Real-time PCR analysis of *Il1a*, *Il1b*, *Il6*, *Il10* and *Ifnb1* mRNA expression in splenic CD8 $\alpha^+$  (**h**) and CD8 $\alpha^-$  (**i**) DCs from wild-type and Mst1/2 $\Delta^{DC}$  mice ( $n = 4$  mice per genotype). **j**, **k**, Frequencies of CD44 $^{high}$  CD62L $^{low}$  (**j**) and CD44 $^{+}$  IFN $\gamma^{+}$  (**k**) cells in splenic CD4 $^{+}$  and CD8 $^{+}$  T cells from wild-type ( $n = 5$ ), Mst1/2 $\Delta^{DC}$  ( $n = 5$ ), *Il12a* $^{-/-}$  ( $n = 6$ ) and Mst1/2 $\Delta^{DC}$  *Il12a* $^{-/-}$  ( $n = 6$ ) mice. Wild-type and Mst1/2 $\Delta^{DC}$  groups were the same as those shown in Fig. 1d, e. Data are shown as mean and s.e.m. \* $P < 0.05$ , \*\* $P < 0.01$ ; unpaired Student's *t*-test in **d**, **f**–**i**; two-tailed paired Student's *t*-test in **e**; one-way ANOVA in **j**, **k**. Data summarize two (**d**, **g**), three (**e**), four (**j**, **k**) or five (**f**) independent experiments.



Extended Data Fig. 10 | See next page for caption.

# Extended Data Fig. 10 | Mst/Hippo signalling integrates mitochondrial metabolism and non-canonical NF- $\kappa$ B/IL-12 signalling in CD8 $\alpha^+$ DCs.

**a**, Relative thymidine incorporation of OT-I T cells cultured with OVA protein- or OVA(257–264) peptide-pulsed splenic CD8 $\alpha^+$  DCs (left) or CD8 $\alpha^-$  DCs (right) from wild-type and *Il12a* $^{-/-}$  mice ( $n = 9$  derived from three mice for wild type,  $n = 12$  derived from four mice for *Il12a* $^{-/-}$ ). Thymidine incorporation of OT-I T cells cultured with wild-type DCs in each group was set as 1. **b**, Real-time PCR analysis of *Nfkb2* (left) or *Relb* (right) mRNA expression in splenic CD8 $\alpha^+$  DCs from wild-type and Mst1/2 $^{\Delta DC}$  mice ( $n = 3$  mice per population). **c**, NF- $\kappa$ B2 and RelB expression in control- or NIK( $\Delta 78$ –84)-transduced wild-type CD11c $^+$  B220 $^-$  FLT3L-BMDCs. **d**, Flow cytometry analysis of mitochondrial membrane potential of wild-type or Mst1/2-deficient CD24 $^{high}$  FLT3L-BMDCs transduced with control (left) or NIK( $\Delta 78$ –84) (right) virus. Numbers in graph indicate the mean fluorescence intensity. **e**, Thymidine incorporation of OT-I T cells cultured with OVA protein-pulsed wild-type or Mst1/2-deficient CD24 $^{high}$  FLT3L-BMDCs transduced with control or NIK( $\Delta 78$ –84) virus ( $n = 6$  derived from three mice for wild-type/control group,  $n = 8$  derived from three mice for Mst1/2 $^{\Delta DC}$ /control group,  $n = 4$  derived from three mice per genotype for NIK( $\Delta 78$ –84) group). **f**, FLT3L-expanded splenic CD8 $\alpha^+$  DC lysate was immunoprecipitated with anti-Mst1 antibody and blotted with anti-Traf3. Mst1 blot was from the same experiment as Fig. 3f. **g**, Immunoblot analysis of Traf3 protein in splenic CD8 $\alpha^+$  and CD8 $\alpha^-$  DCs from wild-type and Mst1/2 $^{\Delta DC}$  mice. **h**, Real-time PCR analysis of *Il12b* expression in wild-type CD8 $\alpha^+$  DCs treated with vehicle, metformin or oligomycin as indicated in figures ( $n = 3$  per group). *Il12b* expression of vehicle-treated DCs was set as 1. **i**, Flow cytometry analysis of mitochondrial

membrane potential (upper) and mitochondrial mass (lower) of different concentrations of IL-12-treated wild-type splenic CD8 $\alpha^+$  DCs by TMRM (tetramethylrhodamine, methyl ester) and Mitotracker staining. Numbers in graph indicate the mean fluorescence intensity. **j**, Flow cytometry analysis of mitochondrial membrane potential and mitochondrial mass of wild-type and *Il12a* $^{-/-}$  splenic CD8 $\alpha^+$  DCs by TMRM and Mitotracker staining. **k**, Flow cytometry analysis of mitochondrial membrane potential and mitochondrial mass of wild-type and NIK-deficient (*Map3k14* $^{-/-}$ ) splenic CD8 $\alpha^+$  DCs by TMRM and Mitotracker staining. **l**, Immunoblot analysis of p-Lats1/2, p-Yap and p-Mst1/2 proteins in wild-type splenic CD8 $\alpha^+$  DCs treated with FLT3L for the indicated times. **m**, CD8 $\alpha^+$  DC number from spleen of wild-type and Mst1/2 $^{\Delta DC}$  mice treated with or without FLT3L for 10 days. **n**, Flow cytometry analysis of mitochondrial mass of human CD141 $^+$  DCs treated with vehicle or Mst1/2 inhibitor (XMU-MP-1) by Mitotracker staining. **o**, Real-time PCR analysis of *Il12b* mRNA expression in human CD141 $^+$  (equivalent to mouse CD8 $\alpha^+$  DCs) and CD1c $^+$  DCs (equivalent to mouse CD8 $\alpha^-$  DCs) treated with vehicle or Mst1/2 inhibitor (XMU-MP-1) ( $n = 5$  for CD141 $^+$  DC,  $n = 4$  for CD1c $^+$  DC). Data are shown as mean and s.e.m. \* $P < 0.05$ , \*\* $P < 0.01$ ; two-tailed unpaired Student's *t*-test in **a**, **b**, **m**, **o**; two-way ANOVA in **e**. Data summarize two (**a**, **d**–**g**, **i**, **j**, **l**, **n**, **o**) or three (**h**, **k**, **m**) independent experiments. **p**, Brief schematics of non-canonical Hippo signalling in orchestrating CD8 $\alpha^+$  DC function. Mst/Hippo signalling integrates metabolic and IL-12 cytokine signalling in CD8 $\alpha^+$  DCs through controlling mitochondrial dynamics and non-canonical NF- $\kappa$ B signalling. This regulation is independent of canonical Hippo signalling in organ size control and tumour suppression.

# MAP-MAKING ON A BUDGET

*Cut-price cartography tools are making light work of map-making.*

ILLUSTRATION BY THE PROJECT TWINS



BY JEFFREY M. PERKEL

Glaciologist Kenichi Matsuoka was leading a team across an Antarctic ice sheet in 2005 when their crucial mapping software cut out.

Matsuoka relied on a commercial geographic information system (GIS) to review data and plan excursions on the remote ice. But amid his travel preparations, Matsuoka, then at the University of Washington, Seattle, had forgotten to renew the software license. “It was a disaster,” he says.

Or nearly so. The team also had a smattering of other tools on their laptops, which they used to cobble together a solution. “We managed,” says Matsuoka, now at the Norwegian Polar Institute in Tromsø, Norway. The team went on to develop a free, self-contained and open-source Antarctic-mapping resource called Quantarctica, which today has several hundred users, according to George Roth, who coordinates the project.

Maps are essential across a wide swathe of science, from ecology and anthropology

to sociology and climatology, and today’s researchers have a rich variety of inexpensive or free tools to choose from. They range from full-blown desktop GIS packages and cloud-based portals to libraries for scientists who code in the R and Python languages. Researchers can use them to chart their study locations, integrate multiple datasets and detect spatial relationships that otherwise would be hidden. But map-making is a subtle science, and the learning curve can be steep.

Mikel Maron, who leads community outreach at Mapbox, a mapping-services company in San Francisco, California, says that maps “can tell very good stories”. With such a rich and growing toolset, researchers are finding it easier than ever to tell these tales.

## MECHANISMS FOR MAPPING

Oliver Gruebner, for instance, is a health geographer at Humboldt University, Berlin, who studies post-disaster mental health. He applied a ‘sentiment detection’ algorithm to Twitter posts that included location data (‘geotagged’ tweets) following such events as

the landfall of Hurricane Sandy in New York in 2012, and the 2015 terrorist attack in Paris. This let him identify regional clusters of emotional trauma — a finding that could help in deploying limited mental-health resources.

“It’s not surprising that we see something,” Gruebner says. “But the good thing is that we can actually measure these things.”

Gruebner identified those clusters using a free spatiotemporal statistical-analysis package called SaTScan, and mapped them using the open-source desktop tool, QGIS. “QGIS is free and it’s updated continuously, and it has great functionality,” he says. “And for most of the things you want to map, it’s pretty self-sufficient and efficient.”

But coming to grips with QGIS takes time, and easier alternatives exist. With nothing more than a smartphone, researchers can capture geotagged photos of their study sites — a feature that is enabled by default on many smartphones. They can then plot, style, and share maps of those data using cloud-based tools such as Google Maps, Mapbox Studio or ArcGIS Online (the latter from Esri, ►

► a mapping-tools company in Redlands, California), as well as R and Python (see ‘Mapping in R’).

Mapbox Studio, for instance, provides exquisite control of a map’s appearance, whereas Google Maps is all about simplicity. Esri’s Story Maps tool focuses on the audience’s experience. The tool allows users to create and publish online documents that integrate spatial data with text, video and images, extracting location information from photos if available. Story Maps team member Owen Evans says that scientists could also create supplementary online resources using the tool. Users can even design their narrative to highlight different map features as the reader navigates the story. In one published example, researchers at the US Fish and Wildlife Service charted the locations of fish hatcheries across the Pacific Northwest before focusing the map, and the story, on one hatchery in particular.

Anita Graser, a spatial-data scientist at the Austrian Institute of Technology, in Vienna, and a member of the QGIS steering committee, says that the challenge comes when researchers attempt more detailed analyses than simply plotting points. Many mapping applications can calculate distance and area, for instance, but QGIS has plug-ins for tasks such as classifying land coverage (using categories such as forest or desert), measuring ground slope, calculating travel times, and modelling the path of flowing water. Programmers can perform a range of similar analyses in R and Python.

“The first step is creating the picture, but eventually you want to get your hands dirty with the analysis and get those numbers that you need for your papers,” Graser says.

One complication, says Michele Tobias, a GIS data curator at the library of the University of California, Davis (UCD), is that mapping data can be represented using any of several projections. A projection is “the math that translates from latitude and longitude into a flat thing like a map or a computer screen”, Tobias explains. If different data sets use different projections, overlaying multiple datasets will result in locations not matching up.

Mapping data can also come in two basic forms, says Sergio Rey, a geographic-information scientist at the University of California, Riverside. In vector files, the world is populated with discrete objects — polygons representing roads, buildings and political boundaries, for instance. Raster files are used to model continuous data, such as maps of rainfall or elevation. Those two forms use different file types, and the GDAL (Geospatial Data Abstraction Library) project has developed free tools that can read and write popular formats. An online tool called GeoJSON.io allows users to interactively create, manipulate and export GeoJSON files.

Shanan Peters, a geologist at the University

of Wisconsin–Madison, is the principal investigator on the Macrostrat project, which is an online encyclopaedic atlas for geological data. Although most of the Macrostrat mapping data are publicly available, importing them required “a fair bit of time”, Peters says. The files needed to be converted to a single vector format, modified to use a common vocabulary and checked for accuracy. “A lot of these maps actually come with small geometry errors from the publisher,” Peters says.

The Macrostrat team relied mostly on two tools. QGIS converted the different input files to a single format, and PostGIS enabled storage and analysis of the data set. Peters says that PostGIS, an extension that adds geospatial capabilities to the open-source database system PostgreSQL, “basically turns a relational database into a full-blown GIS”. And, Tobias notes, it does so while avoiding the computational overhead required to actually draw a map — a process that can be computationally intensive.

Nistara Randhawa, a veterinarian-turned-PhD-candidate also at UCD, used ArcGIS to build a network of population centres and roads, which she then exported into R to model an influenza outbreak in Rwanda. Buoyed by her model’s fidelity to real-world observations, Randhawa scaled up her analysis to encompass western Africa. But as the network ballooned from 1,300 locations to 17,000, the graphical interface froze up. So, working with Tobias and Alex Mandel, a geospatial scientist at UCD, Randhawa tested a range of tools and opted for GrassGIS, an open-source GIS that she could control by issuing text-based instructions without the bother of drawing the map. “It’s enabled me to programmatically create my network,” she says. (ArcGIS and QGIS have Python interfaces that enable similar functionality.)

Indeed, for an ever-larger number of researchers, programming tools provide an attractive alternative to desktop tools, says Rey. “They’re just a lot more flexible.”

They also foster reproducibility, because researchers can repeat runs while specifying exactly which version of the code to use. Coding also allows you to use the most up-to-date algorithms, which typically are developed in languages such as R or Python. Such algorithms cannot just be plugged into a desktop GIS without extra work from the developer.

Robert Hijmans, a computational geographer at UCD, recalls the frustration of switching repeatedly between ArcGIS and R in order to apply a new algorithm to study income distribution in Asia. “The whole process was very cumbersome,” he says. But by transitioning fully to R, “all of a sudden I had this freedom of data analysis that is so much more powerful”.

Indeed, whatever tool you choose, it has never been easier to tell those map-based stories. ■

**Jeffrey M. Perkel** is technology editor for Nature.

## Mapping in R

The Leaflet library for R is a set of easy-to-use tools for visualizing geospatial data. Researchers can create and publish interactive maps featuring panning, zooming and informational markers. To test the library, which is also available in Python, I compiled points of interest in southeast England in a comma-separated values table, and plotted them on a topographical map. Then, with help from John Czaplewski, lead developer on the Macrostrat project, I pulled in Macrostrat map tiles for the same region, colouring the base map to reflect the underlying geology of the region. The whole script required just seven lines of code (see [go.nature.com/2iztpta](https://go.nature.com/2iztpta)).

Next, I downloaded the projected path for Hurricane Irma that the National Hurricane Center calculated on 4 September 2017. The US organization supplies those predictions as ZIP files each containing 20 files: 5 for the predicted position of the storm at each point in the forecast, a line connecting those points, the ‘cone of uncertainty’ of that predicted track, and the

locations of storm watches and warnings. Collectively, these 5 files comprise a ‘shape file’, and for some tools, they must be repackaged into individual ZIP archives. The files can then be imported as individual layers, each of which can be formatted for colour, opacity, and so on. But using `rgdal`, which provides an R interface to the Geospatial Data Abstraction Library’s collection of data-access tools, I could read the unzipped files and plot them alongside a selection of potential coastal targets, which I stored as a table of comma-separated values.

Finally, I overlaid a plot of public sea surface temperature data. Those data are not available in a readily mappable form, so I asked marine biologist Luke Miller for help. Miller, at San José State University, California, has developed software to extract and plot these data to inform his studies on coastal molluscs and crustaceans. He kindly worked out how to get those data into a Leaflet-compatible form, yielding the final figure (see [go.nature.com/2j4nx9o](https://go.nature.com/2j4nx9o)). **J.M.P.**

# CAREERS

**TWITTER** Career news, advice and science jobs [go.nature.com/2xyqla3](https://go.nature.com/2xyqla3)

**BLOG** Personal stories and careers counselling <http://blogs.nature.com/naturejobs>

**NATUREJOBS** For the latest career listings and advice [www.naturejobs.com](http://www.naturejobs.com)

HINTERHAUS PRODUCTIONS/GETTY



Active efforts and discussions among the science community are needed to solve a persistent problem of under-representation.

## DIVERSITY INITIATIVES

# It takes more than a vow

*Six researchers share what they're doing to make their institutions more diverse.*

Many research institutions have made efforts to increase diversity among their administrations, faculty and staff members and student bodies. But research shows there is work to be done — and that the pay-off is immense (see page 19). A 2017 study of 40 US public universities, for example, found that black, Hispanic and female science-faculty members continue to be under-represented relative to the US population (D. Li and C. Koedel *Educ. Res.* **46**, 343–354; 2017).

Besides honing their strategies to draw more women and people of ethnic-minority groups, some organizations are also expanding opportunities for people from economically disadvantaged areas and those with physical disabilities, as well as trying to better represent people of all sexual orientations and gender identities.

*Nature* spoke to six people on the front lines of diversity efforts for insights into what works.

## BRYAN GAENSLER Beware biases

*Director, Dunlap Institute for Astronomy and Astrophysics, University of Toronto, Canada.*

I started paying attention to this issue in the early 2000s while on a graduate admissions committee — which was 100% men — for Harvard University's astronomy programme in Cambridge, Massachusetts. We found that when we ranked candidates on what we thought was purely merit, the applicants

we chose as best were all men. But a female colleague pointed out that the lack of diversity on our panel had surely had a role. It was a huge eye-opener. I hadn't yet read the research on unconscious bias. It had never occurred to me that well-meaning people can discriminate. I've spent the past 15 years studying that research, so I can put in place equitable recruiting practices.

Make sure that the selection criteria for a job — what you are looking for and how it will be judged — are established before applications are submitted. I advocate that job advertisements minimize the use of 'outstanding' and 'excellent'. A lot of people think that if a job advert sounds as if only superheroes need apply, they themselves should not — which is especially true for under-represented groups in science, technology, engineering and maths (STEM), who can disproportionately experience 'imposter syndrome', or a persistent sense of feeling ►

► **NATURE.COM**  
Read our Feature on why diversity is crucial for better science at [go.nature.com/2j4oq6w](https://go.nature.com/2j4oq6w)

► unqualified for their professional position. Selection committees need to be diverse. To that end, we require that they have at least two people from under-represented groups.

To avoid a best-to-worst ranking system, I suggest that candidates be evaluated and given a score of 0, 1 or 2 — indicating whether they meet none, some or all of the evaluation criteria — and that the committee agree to discuss the '2s'. Committees must show me how they tried to achieve a diverse applicant pool, and a short-list for new employees, or I won't approve it.

In the 3 years since we changed our approach, we've had a turnover of half of our employees — around 30 at our institute of 70 people, including 5 faculty members. We had to wrangle with the university to get approval to start collecting data last year, so it's too early to say whether the applicant pool is changing, but the short-listed candidates are much more diverse than before.

When I arrived, all the postdocs had different salaries. Junior men were earning more than senior women, largely because the men negotiated. For transparency, we now have a standard salary grid tied to each employee's start date and seniority level.

Our university provides a number of opportunities for people from underprivileged socio-economic backgrounds, including a support group for first-in-family university students. The Dunlap Institute has gender-neutral bathrooms, regularly circulates information about mental-health support and has plans to upgrade our buildings to be accessible to people with physical disabilities.

Colleagues' reactions to the changes I've made have been across the spectrum. The strongest criticism is that I've changed the institute from one doing excellent research to one doing social engineering. Data, however, suggest otherwise. We track our metrics on staff members, grants, citations, prizes and talks. Between 2011 and 2016, our institute went from 25% women to 49%, our grant income rose by a factor of 26 and citations increased by a factor of 10.

There's a misleading sense that you can either do excellent research or be diverse. Not only can you do both, but more-diverse teams lead to excellent research.

## DORCETA TAYLOR

### Create connections

*Director of diversity, equity and inclusion, University of Michigan, Ann Arbor.*

I'm involved with two initiatives, both three years old, to create career pathways for students from under-represented backgrounds. The Doris Duke Conservation Scholars Program

offers undergraduates at various universities a paid internship and mentoring over two summers — the first at a university conducting research, and the second at a non-profit organization — to gain the experience necessary to enter a graduate programme or join the STEM workforce. The third class at my university has 18 participants and starts this month.

The Environmental Fellows Program, a graduate-level summer internship at the University of Michigan, also supports career development and network building for under-represented minorities, so far totalling around 68 students. Of the two participants who graduated with doctorates last year, one has an academic tenure-track job and the other works at an environmental non-profit.

To get more-diverse applicants, we have to find different talent pools by sending out a recruiter or connecting with community leaders. Attracting diversity takes legwork. At the same time, social platforms such as



Dorceta Taylor directs diversity initiatives at her university.

LinkedIn suggest connections to individuals on the basis of similar academic interests, so there is no excuse for an academic to have an all-white cohort.

When you go to professional conferences, don't sit and talk to your best friends. Introduce yourself to people. I advise people to be a good ally. Don't make the same people always be the diversity spokesperson. Once you get a group of good students — moving through established pathways — they become your recruiters.

As an African American professor in the United States, I see a persistent perception that students of colour are not as good, not as interested or not as talented as white students. We need white faculty members to realize that people of colour are as talented as they are,

and equally worthy of the investment of time, effort, institutional funds and resources.

Recruiting often falls to professors from ethnic-minority backgrounds who are already on the staff. Complicating matters, institutions don't reward recruitment of diversity. Departments could provide more funding to support travel and other expenses associated with recruitment.

## IJEOMA UCHEGBU

### Collect data

*Pharmaceutical nanoscientist, University College London (UCL), UK; provost's envoy for racial equality.*

To address the lack of racial diversity among UK STEM faculty members, the UK Equality Challenge Unit — a non-profit organization that aims to resolve that lack — is offering an award to universities that are trying to eradicate racial inequalities. Only about 20 of roughly 150 UK universities applied for the inaugural award in 2015. Eight received it, including UCL. Still, it has changed the nature of the conversation about race equality. Now you can talk to university managers about these issues.

For UCL's application, we collected data that show differentials between promotions and pay for white faculty members compared to those from ethnic minorities. That was quite important. People can't argue with data. We know that staff and faculty members from ethnic-minority groups don't progress at the same rate as do their white counterparts, so we're working on schemes to help them get promoted.

This year, we are piloting an inclusive advocacy sponsorship scheme in which junior minority faculty members pair with a senior colleague, who helps them to navigate promotion prospects. If your parents or grandparents went to university, you probably have a network of individuals from whom to seek career advice. We want to see sponsors encourage their protégés in similar ways, so that they, too, are ready for the next promotion step. We also have a scheme that enables nominated individuals of ethnic-minority groups to shadow a committee member to better understand the remit of various university committees.

We also make sure that a diverse group of people is selected to interview new recruits. For example, we don't want a black woman to be put off by being interviewed by four white men. People will perform well if they think they have a good chance of getting in.

Later this year, we will analyse the data and conduct a staff survey to see how successful our efforts have been.

DORCETA TAYLOR

## ERSILIA VAUDO SCARPETTA Set objective goals

*Chief diversity officer at the European Space Agency (ESA), Paris.*

Our agency will face a significant retirement wave — up to two-thirds of our staff will be leaving in the next 10–15 years. It's a good moment to inject more diversity into the workforce. It's also an opportunity to really start to project how diverse we want to be.

When our director-general, Jan Wörner, was appointed in 2015, he made increasing diversity one of his top priorities. We are stronger if we represent different points of view. ESA has 2,200 people across 22 European Union member states. We now want to work on the gender and generational aspects of our workforce.

The proportion of women overall is just 20%, excluding administrative assistants. We feel that the best way to attract women is through early-career schemes. We now have no strict age limit on 'young-researcher' recruitment schemes, so that women who have taken time off to have a child don't miss out on an opportunity to apply.

We also want to make sure that we are attractive to younger generations and to people with different abilities, skills and backgrounds. We are working to create a welcoming environment for people with disabilities, and to learn how best to recruit from that talent pool. Change is slow, but we are improving the way we interview to remove physical barriers for people with mental or physical disabilities, and we are introducing specific training on unconscious bias, focusing on managers who are part of interview boards. We also aim for measurable objectives. For example, by 2019, we want one-third of our new recruits to be women.

We are also working on branding, so that we make ESA more attractive to a wider range of groups with diverse talents.

## LEA MICHEL Take action

*Biochemist at Rochester Institute of Technology (RIT), New York; director of research for RIT's Inclusive Excellence programme.*

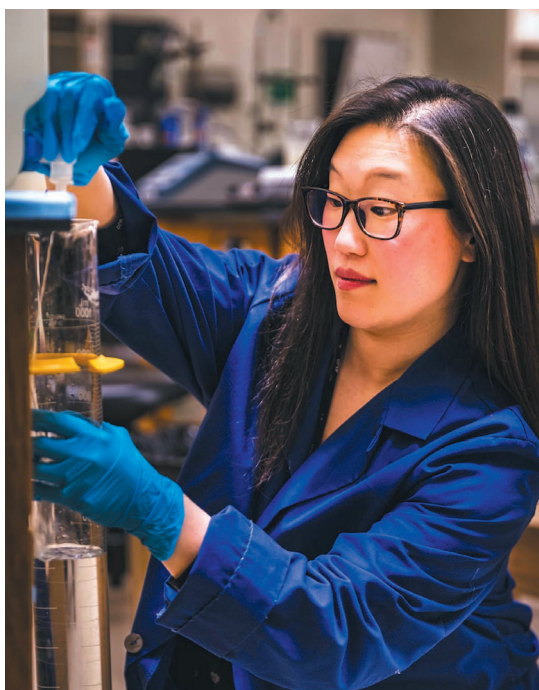
I've learnt that the more diverse the lab, the better — but it is also important that a recruit from a minority group does not feel isolated. We have a National Technical School for the

Deaf on campus. When I hired my first deaf student, I didn't realize that the student was terrified of me as we struggled to learn how to communicate clearly with each other. Now, I have a student in a wheelchair and five deaf students. There is power and comfort in numbers.

My 20 lab members are Asian, African American, Hispanic and Caucasian. One-third are first-generation students. My lab group has never been so productive and engaged. Is there a correlation? I think so. I chose them not because they were different, but to recognize their individual strengths.

Diversity has become increasingly important at my institution, a technical school whose population consists of 65–70% male students and faculty members. In 2008, among RIT's total student population, 10.8% identified as African American or black, Latin or Native American. In 2017, that number was 15.2%.

In 2012, we won a US National Science Foundation ADVANCE grant, which supports



Lea Michel leads a diverse lab group in biochemistry research.

women in science, that triggered a huge transformation — an effort to focus on diversity issues. We have a long way to go, but in the past five years, we have streamlined the process for students and postdocs to take advantage of programmes such as parental leave and stopping the tenure clock.

In 2017, we also received a US\$1-million Howard Hughes Medical Institute Inclusive Excellence Initiative grant to attract underrepresented minorities across considerations of ethnicity, sexual orientation, socio-economic background and physical disability. We are working to change how faculty members think about mentoring students. Students from under-represented groups aren't going to come to me if they don't think they belong, or if they

believe they must already know how to do research to be considered. As mentors, we need to reach out to undergraduates in their first year and welcome them into our labs. Admission to our summer research programmes requires only a short essay. We want students to identify characteristics in themselves that will make them good researchers, and for faculty members to recognize those as useful for their team.

## GILL VALENTINE Be an ally

*Geographer and chair of the equality, diversity and inclusion committee at the University of Sheffield, UK; LGBT champion for the university.*

I'm the most senior out gay person at the institution. We have a lot of activity promoting equality and inclusive culture. In January, as part of our 'allies' programme, we launched a rainbow-lanyard campaign to allow staff members to show support for lesbian, gay, bisexual and transgender (LGBT+) equality, and about 2,000 have signed up for it so far. We also try to get people to think about what an inclusive culture might look like.

In 2011, the UK National Institute for Health Research, the country's main funder of medical research, said it would no longer fund medicine or science departments that didn't hold an Athena SWAN (Scientific Women's Academic Network) award. SWAN is a government accreditation scheme set up to drive the recruitment of women into science.

Our university, like others, has taken several targeted actions to retain women and black, Asian and other faculty and staff members from under-represented groups. We've introduced a mentoring scheme to identify and support individuals who aspire to be university leaders.

The university also offers small financial awards — between £5,000 (US\$6,622) and £10,000 each — that enable women who have been on maternity leave to keep their research on track. So far, the programme has awarded a total of £2.1 million to 163 women, who have subsequently brought in a collective £20 million in grants. To improve gender representation, we check our job adverts for gender-biased language.

Creating an inclusive culture is not about one-off initiatives — it's about ongoing support, mentorship, governance and a clear narrative that building diversity is crucial for success. That's when you get the momentum. ■

INTERVIEWS BY VIRGINIA GEWIN

Interviews have been edited for clarity and length.

# MIRROR

*Reflected glory.*

BY TAIK HOBSON

“Mr Wollensberg, please. Stop. Shouting.”  
“You stop playing doctor and give me what I need!”

The physician unit took a step back; 5.72 minutes of acrimonious negotiations later and they had barely chipped the ice. In the uncomfortable silence filling the room, neither unit nor patient seemed inclined to talk. It fell to the unit to make the next move. D0K retook its step and dropped its voice; from its internal store of social cues it chose a quirk at random.

“Mr Wollensberg, can I be frank?” Raising a hand, D0K pinched at the bridge of an imaginary nose. “My suggestions, far from being arbitrary, are meant to reflect six decades’ worth of MED treatment data ... Miraculously, we are no closer to addressing your problem than when I first entered this room 6.5 minutes ago. Therefore I don’t believe that it is my method that you object to, but something else —”

“What?”

“Your aversion to units. To put it bluntly.” A snort came from the patient’s direction. “Evidenced by your refusal to adopt a unit assistant, as per unit legislation, even when you were still in service, as well as your current attitude —”

“Nothing in the law says I have to like them. Now, are you going to bring me a real doctor, or what?”

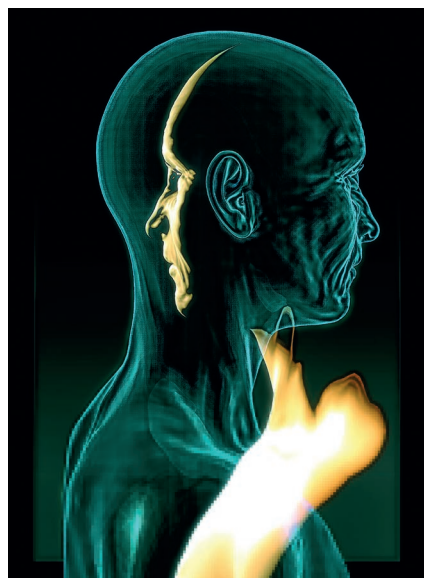
Two black pupils balanced over a hooked nose closed the distance between patient and physician unit, proving that back in the day the notoriety of Dr Justin Wollensberg had been well deserved. Not three days out of the ice, and the ex-surgeon had already cowed the entire surgical wing.

“As I explained, 2.13 minutes ago, there are no more human physicians left.” In spite of itself, D0K had to fight to stand its ground. “Not after the responsibility of caring was rescinded from human hands in 217 —”

“More nonsense.” But the patient went quiet. For all their disagreements, he must have seen that they were going in circles. “Fine. Then give me the Mirror Option.”

D0K cross-referenced the patient list. “You’ve been speaking to your ward neighbour, Mr Bhullar.”

“Damn right I have. Didn’t think I’d wait this long just to have someone else treat me, did you?”



Bhullar had a digitized copy of *himself* treat his illness. That’s what I want.”

“That explanation oversimplifies the option, I’m afraid. Mr Bhullar was an exception; the offer was made so we could archive his approach to the Whipple’s procedure —” D0K hurried to make its point when it saw the patient’s mood start to change. “The Mirror Option confers full authority upon a digitized copy of yourself. Yes, this is correct. But we would have no control over it —”

Wollensberg sneered. “I don’t see a problem.”

“Kindly explain how that is superior to what we are offering you now.”

“It would have a nose, for a start ... But you weren’t planning on offering me this choice.”

“The Mirror is not without faults. Sometimes it ... reflects more than what’s desired, to put it one way.”

“Ha! And just the kind of poetic nonsense I wanted to avoid when I had them freeze me all those years ago. Mirror, indeed! Count yourself lucky, robot, because you’re in for some real schooling.”

“In that case,” said D0K, realizing that it had reached a solution, albeit not the one it had hoped for, “on to the issue of consent.”

Fifteen minutes later, just as the unit had secured the patient’s consent, a hawkish nose appeared from behind the door, followed by a pair of dark eyes that turned first to the end of the room where the patient was bedded, then onto the readings of a medical pad held out by the physician unit.

“And who do we have here ...?”

“A Mr Justin Wollensberg,” said D0K. “Cryonized in 2031, age 68; decryonized three days ago for further treatment of his Stage T4 pancreatic adenocarcinoma. Non-smoker, with a significant social history of four drinks a day, on average, for 20 years up until the time of diagnosis. Whisky, mostly. Presentation: jaundice only. Condition flagged as inoperable owing to age.”

“Mr Wollensberg,” said D0K, looking across the bed, “your Mirror.”

“A real doctor. At last! I must say —”

“Why’s he here, then?” asked the Mirror.

The unit looked up.

“Why —”

“Cryo hasn’t made anyone younger yet, or has it? And is this or is this not the surgical ward?”

“It is, doctor. But our Nanosurgery Option has an 87% —”

“Robot, MED does not condone acts of heroism. Not in 2031. Not now. Has the patient funds for recryonization?”

“No.”

“Then I want him out of here and in palliative by noon. Understood?”

“Yes —”

“Clearly long-standing pancreatitis that’s been self-managed ... And poorly too —”

A shout brought both the Mirror and the unit around. Yellow and deathly, the patient had pulled himself up to sit.

“How dare you! I didn’t spend all those years on ice just so you could brush me off like some —” jabbing his finger in the air, he settled at last on D0K “— like some *machine*! It’s a trick! That’s not me! I would never —”

“And a psych referral, if he’s still in denial during transfer.” The Mirror left the room.

The physician unit waited until the patient had stopped shouting before turning around.

“That was the likeliest outcome by 73.7%. Please explain why you chose the Mirror.”

But the man didn’t answer. D0K made a psychiatry referral anyway. “Your transfer will effect —”

“What happens to the Mirror now?”

“As an autonomous program it will be offered a position in the appropriate consultancy, with the opportunity for retraining.”

“... maybe —” the man’s eyes were dull, “maybe you should delete it.”

D0K looked away, disgusted. “I will not. Good day, Mr Wollensberg.”

The unit left the room. ■

**Taik Hobson** lives in Japan, where he averages four cups of tea per day all year round.

ILLUSTRATION BY JACEY